From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

Statistical models of breast cancer tumour progression for mammography screening data

Linda Abrahamsson

Karolinska Institutet

Stockholm 2018

The front page photograph of four generations of women is taken by Sebastian Abrahamsson in Västerskogen, Sweden 2018. The figures are produced by the author, if not otherwise stated.

Published by Karolinska Institutet.
Printed by E-Print AB 2018.

ISBN 978-91-7831-127-9

Institutionen för Medicinsk Epidemiologi och Biostatistik

Statistical models of breast cancer tumour progression for mammography screening data

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i hörsal Atrium, Nobels väg 12 B, Karolinska Institutet, Solna

Torsdagen den 11 oktober 2018, kl 09.00

av

Linda Abrahamsson

Huvudhandledare:
Docent Keith Humphreys
Karolinska Institutet
Inst. för Medicinsk Epidemiologi och Biostatistik

Bihandledare:
Professor Kamila Czene
Karolinska Institutet
Inst. för Medicinsk Epidemiologi och Biostatistik

Professor Per Hall
Karolinska Institutet
Inst. för Medicinsk Epidemiologi och Biostatistik

Docent Mark Clements
Karolinska Institutet
Inst. för Medicinsk Epidemiologi och Biostatistik

Fakultetsopponent:
Doktor Harald Weedon-Fekjær
Oslo University Hospital
Oslo Centre for Biostatistics & Epidemiology

Betygsnämnd:
Docent Nicola Orsini
Karolinska Institutet
Institutionen för Folkhälsovetenskap

Docent Krista Fischer
University of Tartu
Institute of Genomics

Professor Lars Holmberg
King's College, London
Department of Cancer Epidemiology & Population Health
samt
Uppsala Universitet
Inst. för Kirurgiska Vetenskaper

Stockholm 2018

Till min familj

Abstract

In this thesis we propose a novel statistical natural history model and illustrate how it can be applied to epidemiological breast cancer screening data to increase knowledge about how breast cancers progress over time and how likely they are to be detected by both screening and by symptoms. The model may be useful in helping to design future individualised screening programmes for breast cancer.

In **Study I** a continuous tumour growth model for jointly estimating tumour growth, time to symptomatic detection and mammography screening sensitivity as a function of percentage mammographic density, PD, is presented. The model is applied to data extracted from Swedish postmenopausal breast cancer cases (the same study base is used in Studies I-III). PD is significantly associated with screening sensitivity. Growth rates are found to have a high individual-to-individual variability.

In **Study II** the continuous tumour growth model is extended to allow for covariates in all submodels (tumour growth, symptomatic detection and screening sensitivity). A previously described positive association between body size and tumour size is found to be mainly caused by difficulties in symptomatic detectability/delay in visiting health care.

In **Study III** we compare the statistical powers of detecting image markers related to masking between the continuous tumour growth model and logistic regression using interval vs. screen-detected cancer as the dependent variable. Based on simulated data, we show that statistical power can be higher when tests are based on the continuous tumour growth model. Using observational data, we study an image marker of scatteredness of mammographically dense tissues in terms of screening sensitivity. PD did not include any additional information regarding sensitivity once SI's role in sensitivity was accounted for.

In **Study IV**, using our continuous tumour growth model framework, we derive individual (conditional) lead time distributions, based on a woman's tumour size, screening history and percentage mammographic density. We propose a lead time bias correction that can be used in survival comparisons between e.g. screen-detected and interval cases. In a simulation study, we explore the length-biased sampling. Results showed that the sampling should be viewed in the light of the tumour growth rate *and* the tumour size at which the tumour would have become symptomatically detected in absence of screening.

List of scientific papers

I. Linda Abrahamsson and Keith Humphreys. A statistical model of breast cancer tumour growth with estimation of screening sensitivity as a function of mammographic density. *Statistical Methods in Medical Research* 2016; **25**: 1620–1337. Epub 2013.

II. Linda Abrahamsson, Kamila Czene, Per Hall and Keith Humphreys. Breast cancer tumour growth modelling for studying the association of body size with tumour growth rate and symptomatic detection using case-control data. *Breast Cancer Research* 2015; **17**: 116.

III. Linda Abrahamsson, Maya Alsheh Ali, Kamila Czene, Gabriel Isheden and Keith Humphreys. Using continuous tumour growth models to identify mammography image markers associated with screening sensitivity – unscattered dense tissue masks tumours. (*Manuscript*)

IV. Linda Abrahamsson, Gabriel Isheden, Kamila Czene and Keith Humphreys. Continuous tumour growth models and insights into breast cancer survival and screening. (*Submitted*)

These articles are referred to by their roman numerals throughout, and are presented in full at the end of this thesis.

The studies in this thesis were supported by Vetenskapsrådet, Cancerfonden and the Swedish e-Science Research Centre.

Contents

List of abbreviations and mathematical notations

AD	Absolute (amount of) mammographic density
AJCC	American Joint Committee on Cancer
BFGS	Broyden–Fletcher–Goldfarb–Shanno (a numerical optimisation algorithm)
BI-RADS	A four category (A-D) qualitative classification of mammographic density
BMI	Body Mass Index
BRCA1/2	Breast Cancer Susceptibility Gene 1 or 2
CAHRES	Cancer and Hormone Replacement Study
CISNET	Cancer Intervention and Surveillance Modeling Network
ER	Estrogen Receptor
HER2	Human Epidermal growth factor Receptor 2
HRT	Hormone Replacement Therapy
L	Random variable for lead time
$L(\cdot)$	Likelihood function
$\log(\cdot)$	Base-e log (or the natural logarithm)
MD	Mammographic density
MLE	Maximum likelihood estimation
MLO	Mediolateral oblique view
PD	Percent mammographic density
PR	Progesterone Receptor
R	Random variable for inverse tumour growth rate
$S(\cdot)$	Mammography screening sensitivity function
SI	Skewness of the intensity gradient (measure of scatteredness of MD)
RCC	Regional Cancer Centre
TBA	Total breast area measured in pixels on a mammogram
$T_{det}(\cdot)$	Time in years between tumour onset and detection, in absence of screening

TNM	Classification system used for staging of breast cancers
$V(\cdot)$	Tumour volume function
$V_{det}(\cdot)$	Symptomatic tumour volume function

1 Introduction

To lower mortality from breast cancer, mammography screening is used with the purpose of detecting cancers early in their disease progression, when their curability is high. The age-based screening programmes used in today's health care systems may, in the future, be replaced by individualised, risk-based, screening programmes. In designing such programmes, knowledge of the natural history of breast cancer will be useful. Screening programmes will need to use information on which women have a high risk of getting the disease, but can also make great use of information, at an individual level, on how long breast cancers are likely to be present in women's bodies before symptoms evolve and on the sensitivity of screening.

Tumour characteristics of breast cancer cases are typically only observable at their time points of diagnosis. Still, by adding longitudinal data on screening history, insights about the inherently unobservable processes of cancer progression can be provided, by analysing such data using statistical (natural history) models.

2 Background

In 2012, around the globe, 8.2 million persons died due to cancer. Half a million of these deaths were due to breast cancer, which accounts for 15% of the cancer deaths in women [1]. Advancements in cancer treatment and the implementation of national screening programmes in many countries have made many cancer types less deadly. Cancer cells arise from mutations and lose characteristics that normal cells possess: they resist cell death, sustain proliferative signalling and activate invasion to nearby cells/tissues. These are three of the ten hallmarks of cancer presented by Hanahan and Weinberg [2]. Cancer may evolve in different parts of the human body. Depending on the specific type of cancer, how far it has progressed, the age and any possible comorbidity of the patient as well as where in the world she or he lives, the curability of the disease differs. If cancer is left untreated or is not curable, it may in the later stage of the disease lead to that healthy body organs cease to function, due to loss of space and nutrition demanded by the cancer cells [3].

2.1 Breast cancer

Breast cancer is the most commonly diagnosed cancer among women in the world and especially high incidence rates can be seen in, for example, Northern America, Western Europe, Oceania and Argentina, see Figure 2.1 [1]. A high incidence may be induced by genetic predisposition, a high prevalence of environmental risk factors for breast cancer, as well as overdiagnosis due to screening for the disease. Many of the countries in the above mentioned regions have implemented national screening programmes [4]. The highest age-standardised mortalities are more heterogeneously scattered over the globe and are seen, for example, in such countries as Denmark, Ireland, Ukraine, Pakistan, Afghanistan, some Middle-Eastern countries, Africa's horn, Egypt, Chad, Nigeria, Argentina and Malaysia, see Figure 2.2. In less developed countries breast cancer is the most common cause of cancer death in women, whereas in more developed countries it is the second most common cause, after lung cancer [1].

In the Nordic countries (Sweden, Norway, Denmark, Finland and Iceland) incidence has steadily increased over the last decades [5]. All of these countries have implemented nation-wide screening programmes: Sweden gradually between 1976-1997 [6], Finland in 1987 [7], Iceland in 1987 [8], Denmark gradually between 1991-2010 [9, 10] and Norway gradually between 1996-2005 [11]. Although incidence has increased in these countries, rates of mortality have been decreasing gradually, see Figure 2.3. Annually in Sweden, around 7,500 women are diagnosed with breast cancer and around 1,400 die from the disease [12]. The relative 5-year survival has improved over the last decades, see Figure 2.3 (note that estimates may be influenced by lead time and overdiagnosis, see

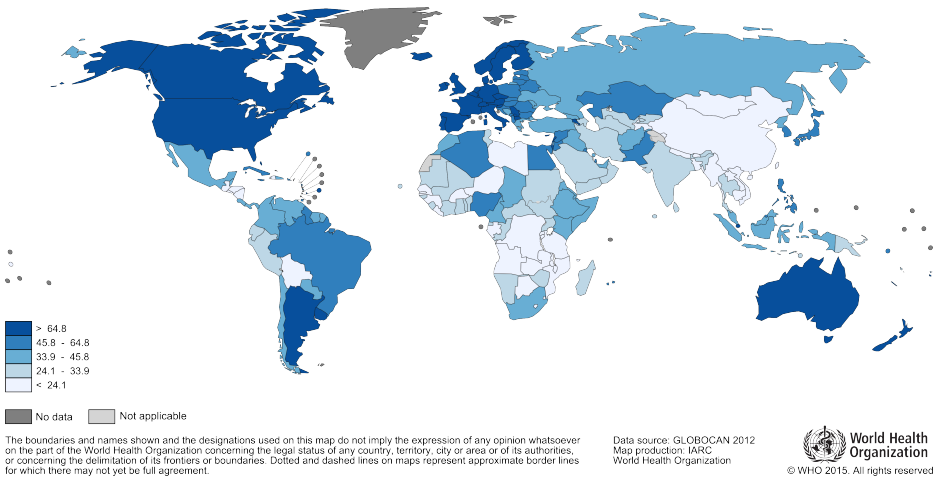


Figure 2.1: Worldwide age-standardised incidence per 100,000 person years. *Source*: GLOBOCAN [1].

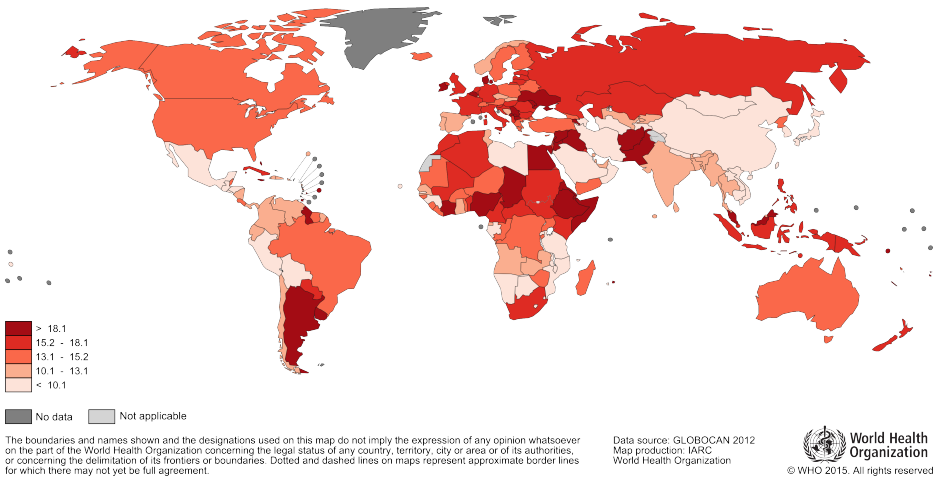
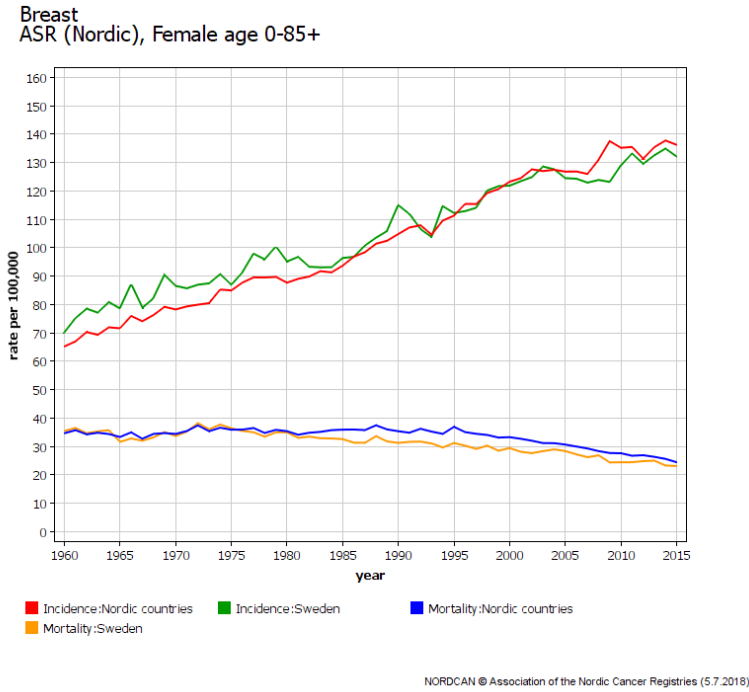


Figure 2.2: Worldwide age-standardised mortality per 100,000 person years. *Source*: GLOBOCAN [1].



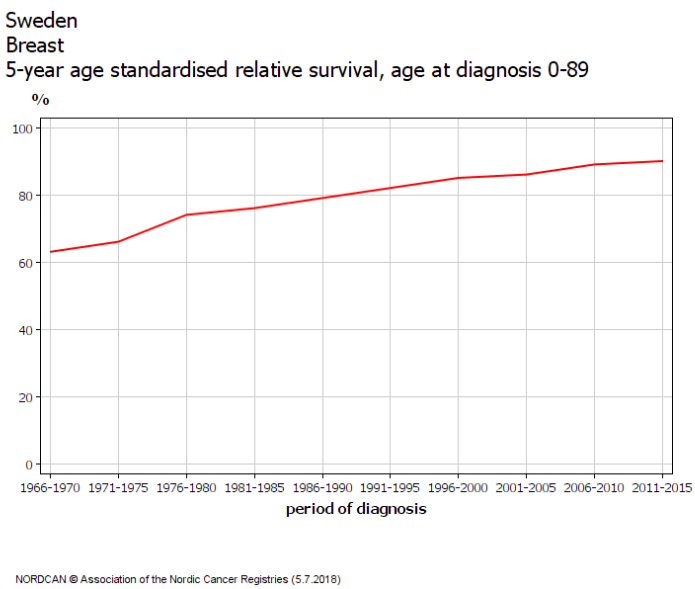


Figure 2.3: Above: Time trends in age-standardised incidence and mortality per 100,000 person years in the Nordic countries. Below: Time trend in 5-year age-standardised relative breast cancer survival in Sweden. *Source*: NORDCAN [5].

Section 2.2.4). The relative survival is the ratio of the observed survival in the patient group to the expected survival in an exchangeable group of the general population (with respect to sex, age and calendar time at the time of diagnosis) [5].

2.1.1 Disease progression

The female breast consists of adipose tissues (fat) and fibroglandular parenchyma. The latter is a collection name of glandular elements and connective tissues (supporting stroma) [13]. The glandular elements contain the lobules and ducts where the milk is produced and drained into the nipple area during periods of breast-feeding [14]. The fraction of adipose tissues to fibroglandular parenchyma varies with age, in such a way that older women have more adipose tissues [15]. Breast composition also changes at times of pregnancy, breast-feeding and during menopause.

The glands are the origin for most breast cancer tumours, which arise more frequently in the ducts than the lobules [16]. Breast cancers can be invasive (invading tissues outside the ducts or lobules) or non-invasive (staying within the gland). There are also benign tumours consisting of cells that lack some of the characteristics of cancer cells (e.g. ability to metastasise), but these are not called cancer. How different the cancer cells are from normal cells is measured by the *grade* of the tumour. Cells with a low grade are more like normal cells and grow slowly, cells with a high grade are less like normal cells and grow fastly [17].

Tumour stage

Over time breast cancers progress through stages. In the *local stage* the cancer is a tumour within the breast that increases in size. After some time, the cancer cells spread through the lymph system to the lymph nodes and the cancer reaches the *regional stage*. In the *distant stage* the cancer cells spread through the bloodstream to distant organs, e.g. the bones, lungs, brain or liver. This is a simplistic explanation, which not all cancers follow. For example, not all cancers may have the ability to become distant. However, when tumour stage is mentioned in this thesis, we refer to this process.

At the clinics, the staging procedure used by physicians for choosing an appropriate treatment, is much more detailed. The classical *anatomic* staging procedure, is based on the TNM classification, in which **T** stands for tumour size, **N** stands for lymph node metastasis and **M** stands for distant metastasis. The three classifiers are each given a score, which are summarised to form a complete breast cancer stage from 0-IV (which also includes subgroups) [18].

In the latest (8^{th}) tumour staging guidelines, the American Joint Committee on Cancer, AJCC, advocate the use of *prognostic* stages, where the breast cancers are classified according to their predicted survival times, such that a higher stage means

worse prognosis. This staging procedure allows for more disease heterogeneity, by the inclusion of factors such as grade, hormone receptor positive status (ER/PR/HER2) and multigene panels [19].

In Norway the (anatomic) stage specific 5-year relative survival, for cases diagnosed between the years 2007-2011, was 100% for Stage I, 92% for Stage II, 76% for Stage III and 26% for Stage IV [20]. This thesis is centered around statistical modelling of the first component (T) in the TNM classification. In data from the Netherlands for cases diagnosed between the years 2006-2012, 5-year relative survival was reported to be 98-101% for T1a-T1c (diameter $\leq$ 2 cm), 92% for T2 (2 cm < diameter $\leq$ 5 cm), 81% for T3 (diameter > 5 cm) and 59% for T4 (tumour growing into chest wall or skin) [21].

Measures of proliferation and rates of tumour progression

At the time of breast cancer diagnosis, proliferation of the cancer cells (in the primary tumour) may be recorded, which describes how fast the tumour is growing/cancer cells are dividing. Since proliferation may vary through the course of the disease progression, this value only measures the growth at the time point of surgery. There are different ways of measuring proliferation in tumours; two of these are by using the *S-phase fraction* or, more common nowadays, the protein marker *Ki-67*.

The S-phase fraction represents the fraction of cells at flow cytometry that are in the synthetic phase (part of cell cycle where the DNA replication occurs, before cell division begins) [22]. Hedley et al. [23] described the methodology in 1987, based on node-positive breast cancer. The protein marker Ki-67 was identified in 1991 by Gerdes et al. [24] as a potential marker of cell proliferation. It has been found to be highly expressed in proliferating tissues, but rarely seen in resting cells [25]. One downside with the measure is that the Ki-67 expression varies in intensity throughout the cell cycle [26], which increases measurement errors/variability. This issue is however common for other markers of proliferation as well [27].

The expression of Ki-67 correlates with the S-phase fraction (and other measures of proliferation) and a high S-phase fraction or Ki-67 value is a sign of poor prognosis [23, 27]. The expert panel responsible for constructing the new AJCC tumour staging guidelines [18], considered using measures of proliferation, but eventually decided not to incorporate them (at least for now), because of technical issues and problems with reproducibility [28]. Both the S-phase fraction (in a study of ~400 cases [22]) and Ki-67 (in a study of ~1800 cases [29]) have been reported to be positively correlated with tumour size. It is however not clear whether results of the second study were affected by the length bias (explained in Section 2.2.4).

Tumour growth/progression describes how fast the tumour is progressing over time and the process is typically unobserved. Statistical models can be built around an assumption of a constant or non-constant rate throughout the course of the disease, see

Section 2.3.3. Rates of growth/progression are likely to be correlated but not equivalent to rates/measures of proliferation, which can vary over time and across the tumour.

2.1.2 Symptomatic detectability

Breast cancer that is not detected by screening is found via symptoms. The most common symptom is feeling a lump in the breast, but other symptoms may be swelling, skin irritation, pain, redness and nipple discharge or retraction [30]. In a Finnish study by Singh et al. [31], which included more than one million screening visits from the national screening programme (both with and without a resulting cancer diagnosis). It was found that in 25% of the recorded visits, symptoms were also recorded, either by the woman or the radiographer. Among the visits that resulted in an invasive breast cancer diagnosis, 36% of the women had recorded symptoms. However, most of the women with recorded symptoms had no present breast cancer.

It has been hypothesised that symptom onset may be delayed in women with large breasts [32–34]. In this thesis a proxy for breast size is used, called total breast area, TBA, measured in pixels from a mammogram. In Figure 2.4 an example of two different TBA's are shown. All mammograms included in this thesis have been altered to e.g. enhance local contrasts and increase the level of details. In these shown mammograms, it seems reasonable that lumps are easier to detect in the smaller breast, but it would of course also depend on the location of the tumour. In a study of close to 500 screen-detected breast cancer cases, with a negative clinical breast examination one year earlier, predictors of poor clinical breast examination sensitivity were reported to be obesity and young age [35]. The reason for young age may however be due to the higher risk for fast-growing tumours in younger women [36–38].

A delayed symptomatic diagnosis could be caused by both a delayed symptom onset and a delayed visit to health care after symptoms have evolved.

2.1.3 Risk factors

Breast cancer is more common in some groups of persons than others. In Sweden a typical breast cancer case is a middle-aged to older woman, the median age at diagnosis was 66 years in 2016 [12]. Although it is rare, breast cancer can occur already in female teenagers and in men. The overall incidence rate of breast cancer in women increases steadily with age. In the United Kingdom, the age-specific incidence rates per 100,000 person years have been reported to be 30.4 and 478.3 in women between 30-34 and 85-89 years old, respectively [39]. A natural effect of this is that incidence of breast cancer may increase in aging populations.

There are many known risk factors for breast cancer, other than age and sex. Some of these are listed below; the emphasis here is however kept to the variables included

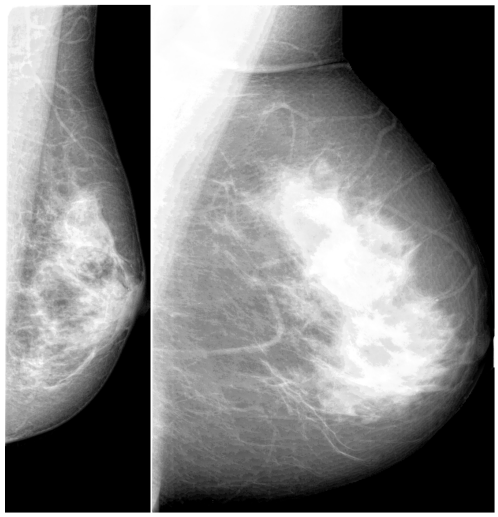


Figure 2.4: An example of a difference in total breast area, with a small breast represented on the left mammogram and a large represented on the right mammogram.

in Studies I-IV. These specific variables have been chosen due to their association with tumour size (body size and breast tissue composition) [32–34, 40, 41].

Breast tissue composition

Tissues in the breast can be categorised as being adipose tissue or fibroglandular parenchyma. The latter tissue type is often called mammographic density, MD, due to its appearance on mammograms, while adipose tissue is called non-dense. This will be described further in Section 2.2.1. MD is a strong risk factor for breast cancer, both in terms of the absolute (amount of) dense tissue, AD [42], and in terms of the fraction of dense tissue, PD [43], seen in a breast, on a mammogram. In a meta-analysis using more than 14,000 cases a relative risk of 4.64 was estimated, comparing women with PD being more than 75% to less than 5% [43]. A high PD has also been shown to be associated with a larger tumour size at diagnosis [40, 41].

Body size

Higher BMI has been found to increase the breast cancer risk in postmenopausal women [44]. The relationship in premenopausal women has been reported to be the opposite, where the BMI at age 18 was even more predictive of breast cancer risk than the BMI at diagnosis [45]. Taller women have been reported to have a higher risk of getting breast

cancer [46]. Reasons behind the link between breast cancer risk and body size are complex [32]. Women with a high BMI are also more likely to have higher amounts of adipose tissues in the breasts, thus BMI is negatively correlated to PD [47]. Body size is positively associated with tumour size at diagnosis [32–34, 48].

Hormone replacement therapy

The use of Hormone replacement therapy, HRT, at menopause is not included as a variable in the models reported in the studies of this thesis. It is however worth briefly describing something about its role in breast cancer here, because of its relation with body size and MD. HRT users have a higher risk of getting breast cancer, which could be due to the exogenous source of circulating hormones [49]. Usage has also been found to increase MD [50], but there is no consistent evidence of weight gain caused by HRT [51]. It has been found that HRT attenuates the effect of body size on breast cancer risk [52].

Other factors

The rare mutations in the *BRCA1* and *BRCA2* genes strongly increase the risk of breast cancer [53]. There exists other more common mutations in other genes, but they do not increase the breast cancer risk to the same extent. Other risk factors related to genes are a recorded *family history* of breast and/or ovarian cancer, as well as *ethnicity*. Some forms of previous *benign breast diseases*, as well as a *previous breast cancer diagnosis* and *high endogenous estrogen or testosterone levels* also increase future risk of breast cancer. Environmental risk factors that increase the risk are *low age at menarche*, *high age at first birth*, *low parity*, *low level of breast-feeding*, *high age at first pregnancy*, *high age at menopause*, *high-dose radiation to chest*, *intake of alcohol*, *current use of oral contraceptives* and *low physical activity* [46]. Hysterectomy, bilateral ovariectomy and mastectomy reduce future risk of breast cancer [54, 55].

A note on menopausal status

In line with the increasing age-specific incidence rates, breast cancer is more common in postmenopausal women than in premenopausal. The causes behind the disease [44], the types of breast cancer [56] and also the breast tissue composition [57] differ somewhat between the groups. Studying both groups in the same analysis may thus lead to difficulties in interpretation of results.

2.2 Mammography screening

In mammography screening low doses of X-rays are used to examine breasts and possibly detect masses that could be cancer. Screening is carried out with the aim that breast

cancers can be detected at an earlier time in the disease progression, when the chance of cure is higher than at the time of symptomatic detection [58]. Some decades ago, randomised clinical trials were performed to find evidence that women invited to regular mammography screening had a reduced breast cancer mortality compared to women that were not invited. Nine trials were included in a recent review (meta-analysis) by the Independent UK panel on breast cancer screening [59]. These trials dated from 1963 to 1991 (in terms of sample size Swedish studies stood for a total weight of close to 50%). Invitation to screening was estimated to reduce mortality by 20% for women aged 50-69 years old. In reviews by the Cochrane Collaboration [60, 61], some of the clinical trials were however criticised for having a suboptimal randomisation procedure. Based on the (three) adequately randomised trials invitation to screening was estimated to reduce mortality by 10%. It however remains unclear whether the effect seen in the trials is similar in clinical settings nowadays, due to better treatments being available [59]. The effect of attending a screening programme on mortality should be higher than the effect of being invited. The effect of attendance is however more complex to evaluate as adherence to screening is not a random event.

As a result of the trials, national screening guidelines and programmes have been (and are still being) implemented in many countries across the globe. In Sweden, recommendations are to invite women from 40 to 74 years old, with an interval of 18-24 months and 70-80% of the invited women participate in the screening programme [62]. There could be a benefit for women also in other age ranges to be screened, although there is no documented evidence of this (clinical trials have not been performed). Women with a genetic predisposition, a family history of the disease or a previous breast cancer diagnosis may follow other routines.

Observational data have also been used to evaluate the impact of screening programmes on mortality reductions. Estimates based on such data are however prone to biases, which is the reason that clinical trials, although old, still stand for the main evidence of the usefulness of mammography screening programmes [59]. Berry et al. [63] presented thorough evaluations of screening in terms of breast cancer mortality in the United States between 1975 to 2000, by summarising seven different statistical modelling techniques performed by different research groups included in the Cancer Intervention and Surveillance Modeling Network (CISNET) Breast Cancer Working Group consortium [64], see Section 2.3.4. Comparing the years 1990 to 2000, breast cancer mortality had decreased by 24% (from 49.7 to 38 women per 100,000). The different research groups came to the conclusion that screening stood for 28 to 65% of the total decrease, the remainder being contributed to by adjuvant treatment (over the years, the use of adjuvant chemotherapy and tamoxifen had been increasing) [63]. In recent work Plevritis et al. [65] presented a similar evaluation but also incorporated information on molecular subtypes of breast cancer.

Benefits from being invited to a mammography screening programme could also be

measured in terms of increased survival instead of reduced mortality. This is rarely done as results would be affected by the lead time bias (Section 2.2.4). It is however clear that, especially at an individual level, survival time as a continuous measure includes more information than mortality as a dichotomous measure.

Whilst the benefits of systematic screening are reduced mortality and prolonged survival, harms can be detection of too many slow-growing cancers (i.e. cases that could be overdiagnosed and consequently over-treated), increased anxiety, use of ionising radiation and economical costs for health care systems [58, 59]. Research is ongoing for evaluating and improving the balance between the benefits and harms of screening programmes, for instance by looking at new ways of individualising screening programmes (not just using the age of women) [66, 67].

2.2.1 Mammography screening sensitivity and mammographic density

A tool used for routine screening programmes should be able to detect the disease of interest, i.e. it should have a high *sensitivity*, so that, in the case of breast cancer, tumours/masses present at the time point of screening are not missed. Mammography is able to detect tumours of very small sizes and it also has the ability to detect changes in breast tissue in the form of *microcalcifications*, which sometimes are a sign of malignant changes [68]. Mammograms are depicted from different views of the breast. The mediolateral oblique (MLO) view is important as it depicts the highest fraction of breast tissue [69]. On mammograms, tumours appear as white masses and are easily seen against a background of fatty tissues (non-dense) which are depicted as dark areas. In low-dense breasts, which are more common in elderly women, sensitivity is high [70]. Even though small-sized tumours can be detected at mammography, it is not, in general, true in mammographically dense breasts. In breasts with a high MD, tumours may be *masked*, since MD also appears white on mammograms. Thus, a high MD both increases breast cancer risk and reduces mammography screening sensitivity.

One qualitative measure of MD is presented in the Breast Imaging Reporting and Data System (BI-RADS) [71]. The measure can be used by radiologists to categorise MD into one of four categories, for assessing risk of masking. In the latest edition from 2013 the categories are constructed as

a) the breasts are almost entirely fatty

b) there are scattered areas of fibroglandular density

c) the breasts are heterogeneously dense, which may obscure small masses

d) the breasts are extremely dense, which lowers the sensitivity of mammography

Other qualitative measures are the mammographic parenchymal pattern developed by Wolfe [72] and the Tabar classifications developed by Gram et al. [73].

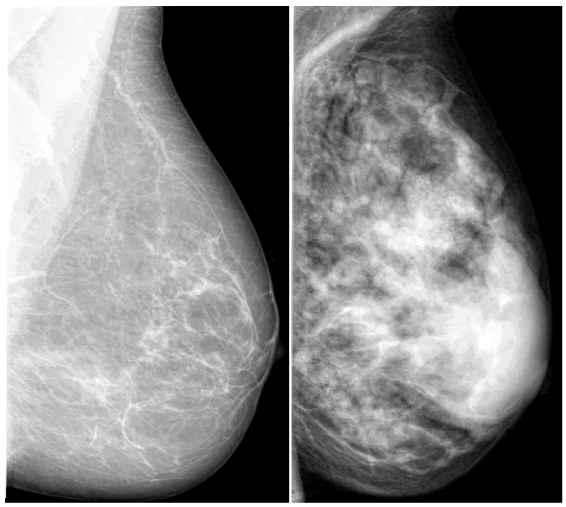


Figure 2.5: An example of a difference in percentage mammographic density, with low PD represented on the left mammogram and high PD represented on the right mammogram.

Quantitative measures of MD are, for instance, absolute mammographic density (AD) and percentage mammographic density (PD). The latter is a function of both the dense and non-dense tissues seen in a breast (the fraction of dense pixels to the total number of dense and non-dense pixels). For creating AD and PD, a threshold for pixel intensity is used for determining whether a pixel is dense or not. Byng et al. [74] developed a semi-automated procedure, called *Cumulus*, where the operator selects the threshold value for pixel intensity on each mammogram. Li et al. [75] developed a fully automated technique, using ImageJ, and validated it against Cumulus. Examples of (two) mammograms with different PD values can be seen in Figure 2.5.

Previous editions of BI-RADS used PD cut-off values for constructing four categories of MD. However, PD (and AD) does not take into account how the MD tissues are divided in the breast in relation to the adipose tissues. Strand et al. [76] found indirect evidence that the image feature *Skewness of the intensity gradient*, SI, developed by Cheddad et al. [77] was related to mammography screening sensitivity. SI is a measure of how scattered the dense tissues in the breast are, i.e. whether the dense tissues are grouped together in large masses or whether they are scattered with streaks of fat between them. An example of two mammograms with different values of SI can be seen in Figure 2.6.

There exist many more image features of MD, which mainly have been assessed for their abilities to determine risk of breast cancer, see for example Zheng et al. [78].

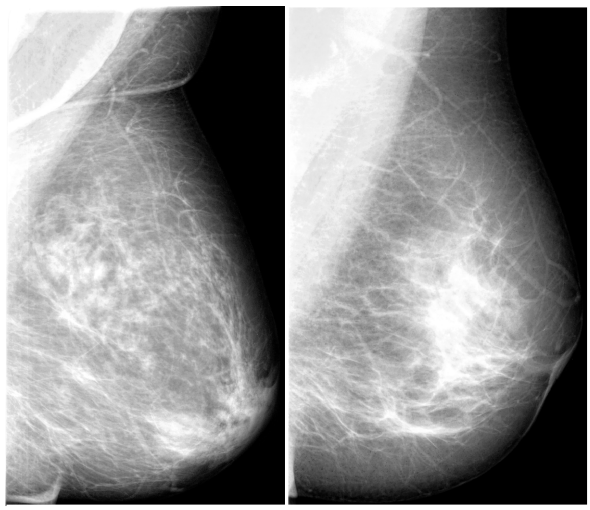


Figure 2.6: An example of a difference in skewness of the intensity gradient, with low skewness (scattered MD) represented on the left mammogram and high skewness (unscattered MD) represented on the right mammogram. The images are from two women with similar percentage mammographic densities (around 20%).

2.2.2 Detection mode and screening history

Before screening programmes (and opportunistic screening) were introduced, detection of breast cancer was done through symptom onset [79]. Nowadays breast cancer can be detected either by symptoms, or by screening (i.e. earlier in the disease progression). How the cancer is detected is usually called the *detection mode* and includes information on, for instance, tumour characteristics. Holm et al. [80] showed that among screen attenders, screen-detected cases had smaller tumours and less lymph nodal involvement than symptomatic cases. Information on *screening history* tells whether a woman has adhered to screening or not. The following descriptions are of types of breast cancer cases, defined in terms of screening attendance and detection mode (based on the ideas of the author), that can be useful to distinguish.

Screening cases

- ◇ Detected at the first (prevalent) screening round
  A cancer that is detected at the first screening round may have existed for a long time in the breast before it was detected by screening.
- ◇ Detected at the second or subsequent (incident) screening round
  A cancer that is detected at a subsequent screening in a woman who attends screening regularly is less likely to have existed for a long time in the breast.

◇ First time to adhere to screening or non-regular screen attender
If a woman suddenly starts to attend screening there may be a reason for this, which is often not recorded. Some screen-detected cases may, for instance, already have felt symptoms.

Symptomatic cases

◇ Not invited to screening
A cancer detected by symptoms in a woman not invited to screening, may be a woman too young or too old to be invited to screening. If such cases are many, there is indirect evidence that the screening programme may need to be extended. Such a symptomatic case can also be found in a place without resources for a regular screening programme.

◇ Not attending screening or non-regular attender
In a Finnish study it was found that women who did not attend screening either went to mammography in other places (false non-attenders), or were more commonly depressed, anxious and socially isolated (true non-attenders) [81]. In true non-attenders, it can be thought that tumour stage at diagnosis may be especially high (if the women do not want to seek health care).

◇ Regular screen attender – **Interval case**
An interval case is a woman adhering to a screening programme, but with a breast cancer symptomatically detected in the interval between two screenings. Understanding the cause of interval cases has been subject to epidemiological research since some of these cancers are very aggressive [80] (having progressed from not detectable at screening to symptomatic within a short time frame). However, this is not true for all interval cases, they can rather be divided into three subgroups (also depicted in Figure 2.7):

- *Masked interval case*
  A marker for this group of cases is that they have high MD and thus a decreased mammographic screening sensitivity [82].
- *Fast-growing interval case*
  Such cases may for instance be seen in women with a low amount of MD, having large tumour sizes at diagnosis.
- *Easy-detectable interval case*
  Such cases may be seen in women with a low amount of MD and small tumour sizes. How gainful screening is for this group of women is not known. Both of the latter groups also exist within the group of women having a high MD.

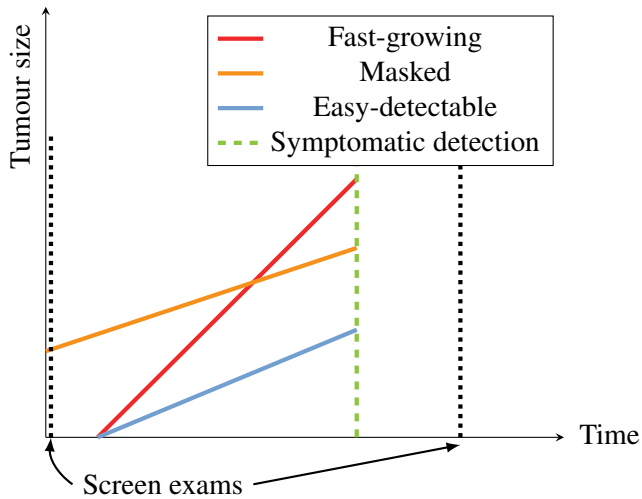


Figure 2.7: Three different types of interval cases. For simplicity, they are all symptomatically detected at the same calendar time point.

2.2.3 Sojourn time, lead time and stage shift

Many interesting factors and processes of the natural history of breast cancer, arising in mammography screening data, are unobservable. In this Section three such latent processes (not already presented) are described.

Sojourn time

Sojourn time is the time from when a tumour is detectable at screening to when it will be detected by symptoms. The concept arises from multi-state Markov models for cancer screening data, see Section 2.3.2. There is however no clear definition of what the term *detectable* means in relation to mammography screening sensitivity and tumour size. The sojourn time is related to the latent processes of tumour growth/progression (how fast the tumour increases in size/the cancer progresses through stages) and symptomatic detectability. In Figure 2.8, two tumours, A and B, are depicted, which have the same sojourn time, but Tumour A grows faster than Tumour B. Factors affecting the sojourn time may be called promoters of clinical disease [83], which should not be thought of as factors affecting only the tumour growth rate.

Lead time and stage shift

Lead time is the difference in time from when a tumour was detected in presence of screening, to when it would have been detected by symptoms in absence of screening.

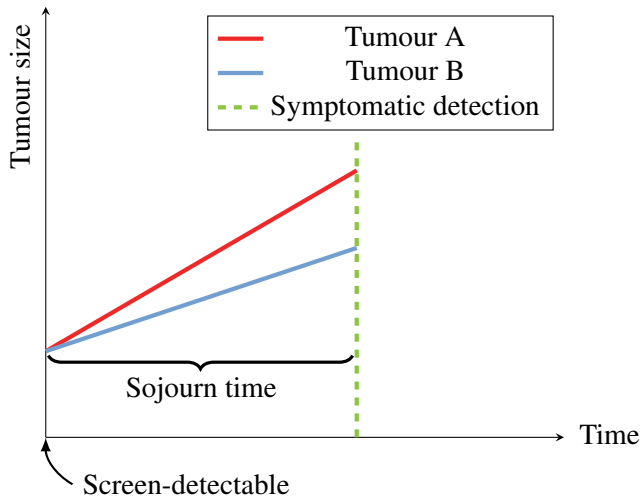


Figure 2.8: Two cases having the same sojourn times, but different growth rates.

In general, for tumours growing with the same speed, a longer lead time gives a greater benefit of screening. However, for a fast-growing tumour, a short lead time may still be of great value. In Figure 2.9, the concept of lead time is depicted. Lead time is related to processes of tumour growth, symptomatic detectability and mammography screening sensitivity.

If we regard tumour stage as a variable that progresses over time, see Section 2.1.1, then stage shift is the difference in tumour stage at the time for screen detection in comparison to that when the symptomatic detection would have occurred, see Figure 2.9. Just as with lead time, the stage shift varies between screen-detected cases. Stage shift and lead time are correlated variables, but screening efficacy will probably depend more on the size of the stage shift (and between which stages the shift appears) rather than the length of lead time.

Survival time from diagnosis for screen-detected cases can be divided into three parts: lead time, survival time that would have occurred in absence of screening and extra survival time due to treatment being started earlier (i.e. the effect of stage shift).

2.2.4 Biases arising in mammography screening data

Comparing breast cancer specific survival, to evaluate the effect of screening, in study populations in the presence or absence of a screening programme or with different detection modes, may be difficult due to biases (true also for comparisons of populations in presence of different kinds of screening programmes). However, not all survival analyses will be biased. Whether or not there will be bias, depends on the underlying

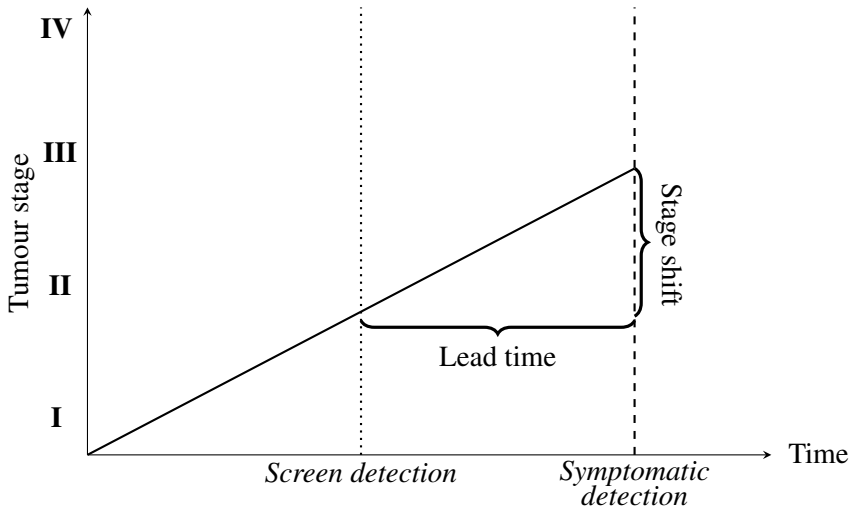


Figure 2.9: An illustration of lead time and stage shift.

question/comparison being addressed. Analyses concerned with the evaluation of independent prognosticators, for data in presence of screening, are also prone to biases [79], as are time trends in breast cancer incidence.

Lead time bias

A lead time bias occurs in comparisons where times to an event are measured from diagnosis in presence of screening and the time point of interest is instead the time for symptomatic detection (in absence of screening). One of the difficulties with this bias is to conclude whether an analysis is affected or not by it, i.e. which detection time point that is of interest to start counting time from. Evaluations of the impact of screening programmes on breast cancer specific survival are subject to lead time bias.

Length bias

Whereas invitation to a screening programme may be randomised, detection mode among screen attenders can never be. Specific groups of women have a higher chance of being screen-detected, thus there exist selection biases, one of these being the length bias. The word *length* comes from the length of the sojourn time, i.e. the time a tumour may be detected by screening but not by symptoms. For a woman with breast cancer attending a screening programme, the probability to be screen-detected increases with the length of her sojourn time. Thus screening cases tend to have longer sojourn times than interval cases, see Figure 2.10. In comparisons of breast cancer specific survival between groups

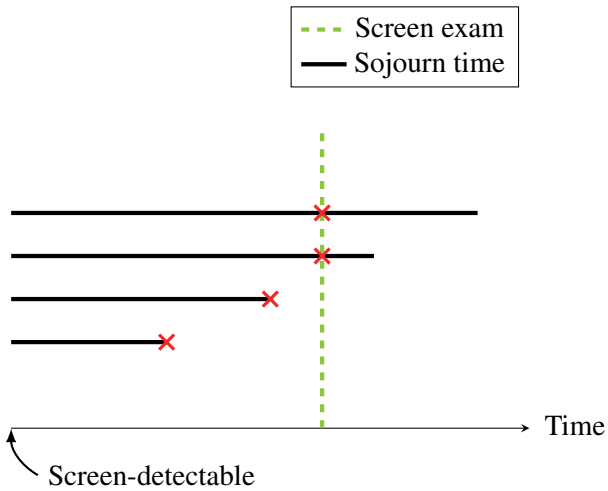


Figure 2.10: An example of length-biased sampling, with four tumours having different sojourn times. The two upper cases will be screen-detected as they were still undetected at the time point of screening. The crosses represent times for detection.

with e.g. different detection modes, any factor, related to the sojourn time as well as to survival (in absence of screening), will infer a length bias. Tumour growth rate is commonly mentioned as such a factor, in the way that screen-detected cases consist of more slow-growing cancers. MD affects the length of sojourn time through the process of mammography screening sensitivity, but as it is probably not independently affecting survival [84], it is not a part of the length bias. The effect of screening on survival is however likely to be smaller in women with a high PD compared to a low, as they are more commonly interval cases and also have larger tumour sizes at screen detection.

Invitation to and attendance of screening programmes

Another form of selection bias is related to the invitation to and attendance of screening programmes [79]. Women attending screening may for instance be more health conscious than the non-attenders, which potentially could cause measures of survival and incidence to differ. Persons newly arrived to another country may not be invited immediately to screening and may not be aware of how the health care system works.

Overdiagnosis

An overdiagnosed case is a screen-detected case, who, in absence of screening, would not have been detected by symptoms during her remaining lifetime. An example is depicted in Figure 2.11. Overdiagnosis is usually considered the main harm of organised

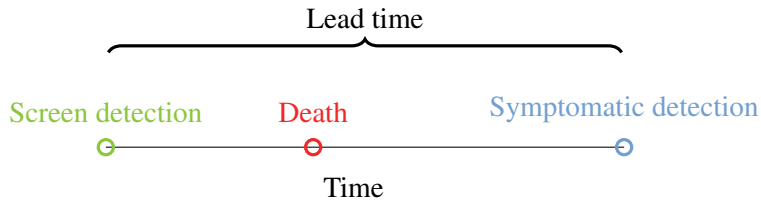


Figure 2.11: An overdiagnosed case where death occurs before the symptomatic detection would have taken place in absence of screening. Lead time exceeds post-diagnostic time to death.

screening programmes, but, unfortunately it is difficult to measure/evaluate. Estimates of the amount of overdiagnosis from breast cancer screening vary from none to quite a large proportion of all breast cancer cases [59]. Typical cases of overdiagnosis are likely to be non-invasive or very slow-growing tumours, or tumours detected in women with other deadly diseases. The amount of breast cancers that are overdiagnosed is likely to be highest in old women. In developing the Dana-Farber CISNET model (see Section 2.3.4), Lee et al. [85] estimated that, for 65-year old women, around 7% of all invasive screen-detected cases will be overdiagnosed, whereas for 75-year old women around 13% will be overdiagnosed. Another harm of screening programmes is overtreatment, which refers to treatment of an overdiagnosed case [86].

2.3 Natural history modelling

Knowledge about how long a tumour exists in the breast before being clinical, when it would have been detected in absence of screening and how sensitive mammography is, are important factors for creating or adapting screening programmes. These are all latent, unobservable processes of breast cancer that cannot be followed directly, for, if nothing else, ethical reasons – once an invasive tumour is detected, surgery and/or treatment will soon begin. The processes may however be estimated by the use of a natural history model of the disease. Data on breast cancer cases and women at risk of getting the disease, collected in presence of screening, provide a rich source of information for unravelling the natural history of breast cancer. One reason being that differences between breast cancer cases in terms of e.g. tumour characteristics become larger in presence than in absence of screening, since detection modes and screening histories vary.

Standard regression techniques are useful for their high generalisability, they may be adapted to many different kinds of data. Related to this thesis work, they have been used to find (positive) associations for tumour size with body size variables and PD [32, 34, 40, 41, 48]. They have also been used to find image features (e.g. PD and SI) related to being an interval or screen-detected case [76, 80]. A common feature of these analyses are

that the mechanisms behind the associations are not directly quantified. Natural history modelling opens up a door for quantifying the causal pathways behind such associations.

In this Section natural history model frameworks and estimation procedures for tumour progression (with a main emphasis on tumour size) and detection are presented. Description of models including joint estimation of survival and tumour onset is beyond the scope of this thesis kappa, and is thus limited.

2.3.1 Study populations used for natural history modelling

For natural history modelling, apart from randomised trials, observational data is often used, in the form of, for example, cohort or case-only studies. Both data in absence and presence of screening may be used, but the processes that are possible to estimate will then differ.

In a cohort study of breast cancer a defined group of women is followed over a specific time period and all new breast cancer diagnoses will be recorded. Such a study has a well-defined risk set (healthy women at risk of getting breast cancer). In a study including only breast cancer cases (e.g. Plevritis et al. [87]) the (full) population at risk of getting cancer is not studied. However, some case-only or case-control studies may be more detailed, including more variables for analysis than registry-based cohort studies. For both types of studies, it is theoretically possible to estimate the natural history processes occurring after tumour onset, such as sojourn time, tumour progression, symptomatic detection and mammography screening sensitivity, under particular (modelling) assumptions. For cohort data it will usually in addition, be possible to estimate incidence rates [88] and evaluate risk factor associations.

2.3.2 Multi-state Markov models

Multi-state Markov models are the most common basis for modelling the natural history of breast cancer and have been applied extensively in the literature, often based on data from randomised screening trials [36, 88–90], but also on observational data [85, 91, 92]. The model assumes that a breast cancer passes through different discrete states in its disease progression. Although the progression is discrete, time may be modelled as continuous. The basic multi-state Markov model is based on the Markov property, i.e. it is a memoryless stochastic process, which infers that the time spent in each state is exponential. It means that the probability for the disease to be in a specific state at a future time point is independent of the track record before the time point t, if the state at time t is known [93]. In further assuming that the process is time homogeneous (does not change with calendar time) it is possible to define a matrix $\mathbf{P}$ of transition probabilities $p_{jk}(t)$ between states, such that $p_{jk}(t) = P(\text{disease in state } j \text{ has transitioned to state } k, t \text{ time units later})$. The matrix $\mathbf{Q}$ is the corresponding matrix of (immediate) transition rates

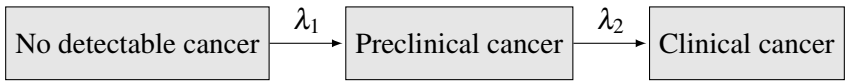


Figure 2.12: A basic state space model for the natural history of breast cancer in presence of screening.

$q_{jk}(t)$ between the states. If the state space is finite and $\mathbf{P}'$ is the matrix of derivatives of the transition probabilities, then $\mathbf{P}' = \mathbf{PQ}$ holds. If another distributional assumption than exponential is made for the time spent in each state, the model is called a Semi-Markov model [90].

The most basic Markov model for breast cancer data is in absence of screening and only has two states: *No detectable cancer* and *Clinical cancer*, thus it includes only one state transition. It can be used to estimate incidence of clinical disease. However, in practice, standard regression techniques (generalised liner models, GLM's) are used for such evaluations. A three-state Markov model that has been applied more often in the research literature is one for data in the presence of screening, which includes one extra state between the two previously mentioned, called *Preclinical cancer*. This is the state when the tumour is detectable at screening but not yet by symptoms and the time spent in this state is called the sojourn time, as previously described (Section 2.2.3). In this model two transition rates can be estimated, which we may call λ_1 and λ_2, see Figure 2.12 [88]. Mammography sensitivity in this model is the probability that a cancer in the preclinical state is detected at screening [36]. Theoretically, the number of states to be included in a multi-state Markov model is limitless, but in practice restricted according to the size and type of data source that is going to be used. For instance, tumour size may be incorporated such that there exists two preclinical states and two clinical states with small and large tumours (similar to the model described by Uhry et al. [91] but for lymph nodal involvement). Then transitions may occur from *No detectable cancer* to *Small preclinical cancer*, from *Small preclinical cancer* to *Large preclinical cancer* or *Small clinical cancer*, and from *Large preclinical cancer* to *Large clinical cancer*. In such model screening sensitivity may be assumed to either differ between the two preclinical states [89], or be the same throughout both states [91].

To estimate the transition rates and screening sensitivity, maximum likelihood estimation (applying varying forms of likelihood functions [92, 94]) and Bayesian methods [89] have, for example, been used. To the best of my knowledge, all of these models require information on women at risk of developing breast cancer (i.e. has not been developed for case-only data). Multi-state Markov models can also be used to evaluate lead time and overdiagnosis [85, 95, 96], but tumour growth rates cannot be directly evaluated, although they are strongly correlated to transition rates between preclinical to clinical states. A commonly used lead time correction procedure for

survival measures, described by Duffy et al. [96], is based on the three-state Markov model. The approach is based on the assumption that lead times for screen-detected cases follow the same exponential distribution as the sojourn times in the entire breast cancer population. The correction procedure takes away the same amount of survival time for all (screen-detected) women (it is unconditional on tumour characteristics, screening history, etc.). It however ensures that no individual lead times exceed times to breast cancer death from times of diagnosis.

2.3.3 Continuous tumour growth models

Both multi-state Markov models and continuous tumour growth models view time as continuous. However, as opposed to multi-state models, which consider cancer progression in terms of discrete events (from state to state), continuous tumour growth models assume that tumours continuously (cell by cell) increase in size. Regional and distant spread may still be described as discrete events. The models require an assumption of how tumours grow over time. In this Section we present continuous tumour growth models that make use of analytical methods for estimating unknown model parameters.

Tumour growth laws

Different laws of tumour growth have been proposed for use in continuous tumour growth models, based on varying kinds of evidence. The most basic tumour growth law is the one assuming that tumours increase in volume exponentially [97], such that

$$\frac{\mathrm{d}V}{\mathrm{d}t} = r_e V(t), \tag{2.1}$$

where V is the tumour volume, t is time and r_e is the growth rate. From this model it follows that the tumour volume doubling time is constant over time ($= \frac{log(2)}{r_e}$). The model implies that the nutrition of the tumour does not cease with time. Among researchers there is a general agreement that the exponential model is a valid assumption for early tumour growth [97].

To allow for a decelerating tumour volume doubling time other growth laws have been suggested. The power law model has the form

$$\frac{\mathrm{d}V}{\mathrm{d}t} = r_p V(t)^{\alpha}, \tag{2.2}$$

where growth is decelerating for $\alpha < 1$ and r_p represents the growth rate. The generalised logistic growth law has the form

$$\frac{\mathrm{d}V}{\mathrm{d}t} = r_l V(t) \left(1 - (V(t)/K)^{\beta}\right), \tag{2.3}$$

where r_l is the growth rate and K and β are constants.

Talkington and Durrett [97] used a small sample of screen-detected breast cancer cases (previously described in Heuser et al. [98]) that could be seen on two mammograms at different time points; at the diagnostic screening and at an earlier negative screening (where the tumour was detected in retrospect). They found that the exponential growth had a superior fit over power law, generalised logistic and Gompertz growth laws for breast cancer tumours.

In previous work, Spratt et al. [99] used a similar, but larger data set of breast cancers with more observations per woman over time. They concluded that tumour growth was best described with a variant of the logistic growth function, together with lognormally distributed tumour growth rates.

In more recent work on breast cancer tumours in mice, Hartung et al. [100] found that breast cancer tumour growth was best described by Bertalanffy, West and Gompertz models, which all involve an initial exponential growth phase.

Continuous tumour growth models in absence of screening

To model tumour size distributions for case-only data in absence of screening Atkinson et al. [101] and Brown et al. [102] proposed to use a combined model of tumour growth and symptomatic tumour detectability. For time to symptomatic detection they proposed the use of a hazard function, dependent on tumour volume raised to some power α. Klein and Bartoszyński [103] found a good fit of the model for $\alpha = 1$ (shown in Section 5.1). For the tumour growth part Atkinson et al. [101] and Brown et al. [102] proposed the use of an exponential tumour growth, with gamma distributed inverse tumour growth rates to account for individual-to-individual variation (shown in Section 5.1). In work on lung cancer, Bartoszyński et al. [104] found that the model fit was superior when assuming inverse tumour growth rates followed a gamma distribution instead of being deterministic. Also, Brown et al. [102] found that the model fit to breast cancer tumour sizes was as good for their proposed model as when assuming Gompertzian or logistic growth laws, which are mathematically more complex.

The proposed submodels by Atkinson et al. [101] and Brown et al. [102], make it possible to derive the tumour volume distribution in absence of screening, $f_{V_{det}}(v)$, as shown by Plevritis et al. [87]. In their work, they also extended the modelling of tumour sizes to include regional and distant stages by using similar hazard functions (as for symptomatic detection) dependent on tumour volume. Parameter estimation was made by maximising a product of three likelihood functions for tumour size, conditional on tumour stage (local, regional or distant) at diagnosis. (In terms of analytical approaches for estimation of parameters, maximum likelihood estimation has been the main procedure used for continuous tumour growth models.) Using the same hazard function for time to symptomatic detection, Chia et al. [105] assumed a more complex tumour

growth model, described by a geometric Brownian motion. However, not all parameters of interest in tumour growth and symptomatic detectability functions can be estimated based on data in absence of screening [87].

Continuous tumour growth models in presence of screening

Continuous tumour growth models have been used sparsely for data in the presence of screening. Weedon-Fekjær et al. [106] proposed the use of a model for tumour growth and mammography screening sensitivity. For tumour growth they assumed the logistic growth model presented by Spratt et al. [99] having the form

$$V(t) = \frac{V_{max}}{\left[1 + \left(\left(\frac{V_{max}}{V_{cell}}\right)^{0.25} - 1\right) e^{-0.25\kappa t}\right]^4}, \tag{2.4}$$

where t is the time, κ is the growth rate assumed to follow a lognormal distribution, V_{cell} is the volume of one cell, V_{max} is the assumed maximum volume (corresponding to a tumour diameter of 128 mm). For mammography screening sensitivity they proposed the use of a logistic form, as a function of the tumour diameter at the time of screening, see Section 5.1.

Based on a well-defined cohort of women attending the Norwegian Breast Cancer Screening Programme the authors presented a likelihood function in two parts to be used for parameter estimation. The likelihood function modelled tumour sizes in screening cases detected at the prevalent screen (without previous screen attendance) and incidence of interval cases. In additional work [92] a likelihood function for tumour size in screening cases detected at screens subsequent to the entry screen (given time since last negative screening) was described. Instead of assuming a model for time to symptomatic detection, as was done by Atkinson et al. [101] and Brown et al. [102], an additional data set of tumour sizes in absence of screening (before the screening programme was implemented) was used. To derive the probability of tumour sizes for screen-detected cases, they applied back calculation from this external data source of symptomatic tumour sizes.

Hanin and Yakovlev [107] had earlier described an (unrelated) approach (which was mathematically technical but not in detail applied to real data) for constructing joint likelihood functions for age and tumour size at detection for data in presence of screening, which is applicable to data from randomised screening trials. They used the previously mentioned tumour growth and symptomatic detection models of Atkinson et al. [101] and Brown et al. [102], together with a discrete conditional hazard rate for time at screen-detection, assumed to be proportional to tumour size (previously described by Hanin [108]).

2.3.4 Simulation-based approaches

As an alternative to using analytic methods to directly estimate parameter values in natural history models, simulation-based procedures with model calibration against real data can be used [109]. Chia et al. [105] suggested that Monte Carlo simulations provide a useful tool in situations where a likelihood function is mathematically difficult to derive. In a Monte Carlo simulation, or microsimulation, for breast cancer in presence of a screening programme, the life-histories of many individuals can be simulated, including events such as birth, screening exams, tumour onset and progression, clinical diagnosis and death due to breast cancer or other causes [110]. These can be simulated under pre-specified model assumptions, which can, for example, be of either nature – multi-state model or continuous tumour growth model. As an example, Tan et al. [110] used Monte Carlo simulations to estimate parameters for a lognormally distributed tumour growth rate (assuming an exponential tumour growth), a Weibull distributed threshold tumour size for screen-detection, and a lognormally distributed tumour size at clinical diagnosis (for detection caused by the size of the primary tumour, distant metastasis was however also a possible cause for symptomatic detection). The authors calibrated their model against breast cancer specific survival stratified on tumour size and detection mode based on data from the Swedish two county study, by repeatedly evaluating simulated histories using the score function method in combination with a quasi-Newton optimization procedure. However, different approaches to calibration are possible. Unfortunately, researchers do not always describe explicitly how calibration was carried out in such publications [109].

The CISNET consortium

Some of the studies mentioned in Section 2.3 have been carried out within the CISNET Breast Cancer Working Group consortium, which was formed in 2000 to use statistical and simulation modelling for evaluating the impact of different breast cancer control strategies on incidence and mortality [64]. In the consortium for breast cancer, six research groups (originally seven) developing six independent models are included. Collectively these groups have produced many publications. The six models are called **Model D**: The Dana Farber Model, **Model E**: The Erasmus MC MISCAN Model, **Model G**: The Georgetown-Einstein Model, **Model M**: The MD Anderson Model, **Model S**: The Stanford Model, **Model W**: The Wisconsin-Harvard Model. Some of the models are based on multi-state (Markov) models for the natural history component and some on continuous tumour growth models. Some estimate parameters using analytical approaches, but mainly simulation-based approaches have been used. A recent highlight presented by the consortium is the comparative study by Plevritis et al. [65], described in Section 2.2, which concludes that the mortality reduction from screening is largest in the group of women having ER- and HER2-receptor negative disease. The probability for

this group to be detected at screening is however in general lower, see also Munoz et al. [111].

Model E is presented in, for example, the already mentioned study by Tan et al. [110], where the natural history parameters estimated by using a simulation-based approach (already described in Section 2.3.4) were used as model inputs in the MISCAN-Fadia model. This was done (also that by using a simulation-based approach) to evaluate the reasons behind the mortality reductions seen in the U.S. population. The natural history component in the MISCAN-Fadia model comprises one of four parts of the complete simulation set up, the others describing the population, mammography screening and adjuvant treatment.

Model D is the only model in CISNET for breast cancer that takes on an analytical approach for estimating parameter values. In a recent development by Lee et al. [85] from 2018, the natural history component (pre-death) consists of a six-state model, having the same states as in a basic three-state Markov model, but also including *preclinical undetectable DCIS*, *preclinical screen-detectable DCIS* and *clinical DCIS with symptoms*. The sojourn time for invasive tumours is assumed to be exponential, but not for DCIS tumours (abbreviation for ductal carcinoma in situ – these tumours are less aggressive than the invasive ones). The authors also present analytical expressions for lead time distributions, to be used for evaluating overdiagnosis – see results presented in Section 2.2.4. The natural history model component does not take tumour size/cancer stage into account. Neither do the lead time and overdiagnosis expressions.

Model S is based on a Monte Carlo simulation procedure, producing life-histories of women, which can be calibrated to U.S. mortality data [112]. For the natural history part of the model, assumptions are similar to the ones used in Plevritis et al. [87], but include a dependency of age and menopausal hormonal treatment on growth rates. Screen-detection occurs if a tumour is larger than a specific threshold size at the time of screening. Estimation of natural history parameters, including a hazard rate for time to symptomatic detection, and one of the two unknown parameters of the inverse tumour growth rate gamma distribution (under the assumption of one parameter being constant), are made analytically based on data in absence of screening [113]. The other growth rate parameter is fitted via a simulation-based approach, calibrated to U.S. incidence data, as is the threshold for tumour size at screen-detection [113]. A new estimation procedure is however being developed [112].

To briefly summarise the modelling of the natural history components in the remaining models: Model G applies a multi-state model, Model W assumes tumours grow exponentially and assumes a model described by Shwartz [114] and Model M applies no natural history component [115, 116]. Overall, in the models, use of mammographic breast density is limited. Half of the models (all of the ones assuming a continuous tumour growth) are based on an exponential tumour growth law.

3 Aims of this thesis

The overall aim of this thesis is to form the basis and show the potential, of a statistical model for the natural history of breast cancer in presence of screening. It is hoped that the model can be of value for future research in designing individual screening programmes. The specific aims of the four studies are:

- ◇ To develop a novel statistical model that jointly estimates breast cancer tumour growth rate, time to symptomatic detection and mammography screening sensitivity as a function of percentage mammographic density.
- ◇ To enable and exemplify the use of covariates in all parts of the proposed continuous tumour growth model and to explain the association between body size and tumour size in terms of breast cancer tumour growth and time to symptomatic detection.
- ◇ To show the strength of the continuous tumour growth model in terms of its ability to identify novel image markers important for mammography screening sensitivity. To quantify the role of scatteredness of dense tissues in mammography screening sensitivity, in relation to percentage mammmographic density.
- ◇ To derive lead time distributions on an individual basis and use these to develop a novel lead time bias correction to be used in survival comparisons between screening and symptomatic cases. To extend the traditional view of length bias.

4 Data material

In this thesis, both observational and simulated data on postmenopausal breast cancer cases have been used. In this Section the use of observational, case-only, data is described, while data simulations are briefly explained in Section 5.6.

Swedish registries and linkage between them

In Sweden there are six Regional Cancer Centres (RCC's) which originate from 1958 and nowadays are responsible for the Swedish Cancer Registry [117]. It is compulsory for physicians and e.g. pathologists, to report all diagnoses of malignant cancers to the registry. The registry includes information on date and basis of diagnosis and cancer invasiveness. It also has a high completeness.

All persons born in Sweden or living in Sweden for at least a year are registered in the Swedish Population Register, held by the Swedish Tax Agency [118]. The register includes information on, for instance, addresses and personal identity numbers (a 10-digit identifier that is unique to each Swedish resident). For medical researchers registries such as these may be linked upon request, through the personal identity number, if ethical permission has been given by an Ethical Review Board.

CAHRES

CAHRES (Cancer Hormone Replacement Epidemiology in Sweden) is a case-control study, which includes cases diagnosed with incident primary invasive breast cancer between October 1993 and March 1995 and age-matched controls. Its original aim was to study the effects of HRT on breast cancer risk [119]. The study included women between 50-74 years old and the study base was determined using the Swedish Cancer Registry, through which 3979 women were found to be eligible cases. Out of these 3345 cases (84%) participated in the study and from most of these women questionnaire data, as well as information from medical records, were available. In extensions of the study, information on screening history (up until five years prior to diagnosis) and mammograms were collected from radiology departments and mammography screening units in Sweden [40]. In this thesis, only cases have been included.

Study populations

While Study IV makes use of simulated data, Studies I-III are (partly) based on observational data, all originating from the same source population, CAHRES. For all

study populations women without written consent, with a previous or other cancer diagnosis (except from non-melanoma skin cancer), with a noninvasive breast cancer or with a diagnosis made outside the study period, were excluded. As breast cancers tend to differ depending on menopausal status, see Section 2.1.3, only the group of postmenopausal women has been studied. Analysing the premenopausal cases separately would lead to low statistical power, due to the original age inclusions of CAHRES. For women with unknown age at menopause, the 90th percentile of 55 years at menopause was used as a cut-off value for inclusion. A flowchart for the study populations is presented in Figure 4.1. Study I consisted of 1370 cases, Study II of 1352 cases and Study III of 1845 cases. This thesis work has been carried out continuously over a long time period and quality checks of the data have also been made continuously, which has resulted in some differences between the study populations. The main difference being the inclusion of more women in Study III due to more mammograms being available.

The main focus of this thesis is on methods development. The number of variables used has been few but these have been chosen with care.

Tumour size

Throughout this thesis breast cancer tumour size provides the main information used for statistical modelling. (The importance of tumour size in relation to e.g. survival has been exemplified in Section 2.1.1.) In the data, tumour size represents the maximum distance of the tumour in mm, measured from pathologists' examinations of tumours removed at surgery. If a woman had more than one tumour in a breast, the largest one was recorded. The variable is continuous, although it is clear from inspection that pathologists tend to approximate tumour sizes to the nearest 5 or 10 mms. For larger tumours, approximations appear to be coarser. Because the use of neoadjuvant chemotherapy, before surgical removal, can downsize tumours [120], women receiving this treatment were excluded in Studies II and III.

Detection mode

How a breast cancer case was detected (by symptoms or at screening) is recorded as a binary variable. Reasons for mammography/diagnosis were retrieved from the records where mammography was performed, as well as from the medical records. Women with conflicting information were excluded in Study III. A further refinement in Study III was that women were excluded if reason for mammography was recorded as being *control* or *no reason*. In Studies I and II, these women were coded as screen-detected cases.

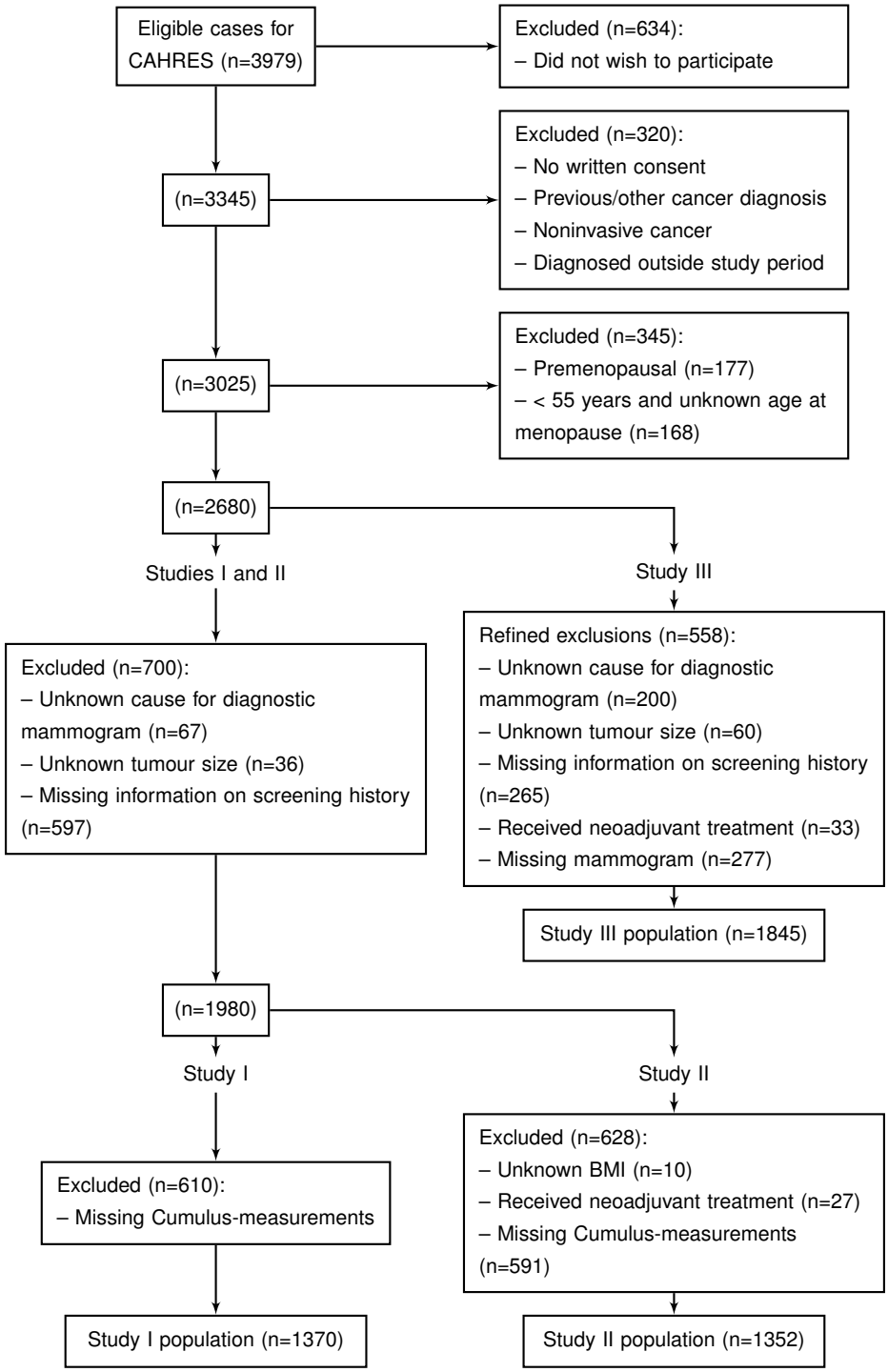


Figure 4.1: Flow chart of study populations in Studies I, II and III.

Screening history

Information on screening history was searched for from five years until three months before diagnosis of breast cancer. The three months cut-off was used to avoid wrongly including any mammography included for the diagnosis work-up, as part of the screening history of negative screenings. In the analysis, a maximum number of three negative screenings were used. In the questionnaire to the cases, women were asked about whether they had been to screening and how many times. We excluded women who answered that they had been to screening *at least once* (Studies I-II) or *more than once* (Study III) the last five years, but had no data on negative screen occasions. The refined exclusion for Study III was made, since it is not clear from the questionnaire that only negative screenings should be considered.

BMI

BMI was calculated from questionnaire data of weight and length of the women at study entrance (post-diagnostic values). The women were also asked about their weight one year before study entrance and the differences in values between the two variables were small.

Percentage mammographic density

Studies I and II

In Studies I and II, we aimed to use the semi-automated "golden-standard" Cumulus-measurements of percentage mammographic density already available from an extended study of CAHRES [40], produced by a trained observer. Thus, one mammogram per woman was selected for inclusion in the studies – the mammogram closest to (before) diagnosis in the contralateral breast, to avoid the appearance of tumours. Only mammograms of mediolateral oblique view (MLO) were considered. However, approximately 600 cases had to be excluded due to missing mammograms or Cumulus-measurements.

Study III

To further increase statistical power we decided to, instead of using the Cumulus-measurements for percentage mammographic density, use the corresponding measurements from an automatised procedure [75]. Such measurements are not time-consuming to provide and thus they could be produced for, close to, all collected mammograms. However, to some recorded screen occasions mammograms had not been retrieved. All contralateral (MLO-view) mammograms available between the years 1988-1997 were included in the study. For each woman the average PD of all her included mammograms were used in the analyses, to reduce measurement errors.

Scatteredness of dense tissues

In Study III, as a proxy for the scatteredness of dense tissues, the technical measure of Skewness of the intensity gradient, described by Cheddad et al. [77], was used. The same set of mammograms as for PD were included for analyses and for each woman the average SI of all her included mammograms was used.

Breast size

We did not have a perfect measure of breast size available for Study II, rather a proxy for it, total breast area, TBA, was used. TBA measures the amount of pixels on a mammogram, which together constitutes the dense and non-dense areas. TBA was available for the same mammograms as PD and was measured by the same observer.

5 A novel continuous tumour growth model

In this thesis, a novel continuous tumour growth model for data in presence of screening is presented. The model consists of three submodels: *tumour growth*, *time to symptomatic detection* and *mammography screening sensitivity*. It is the first continuous tumour growth model that analytically and jointly estimates the three processes, while allowing for inclusion of covariates in all parts of the model. It does not model the onset of tumours and can thus be used for case-only data (i.e. does not need data on the complete set of at-risk individuals). The model has a closer link to tumour biology, than the multi-state Markov models presented in Section 2.3.2, which use discrete tumour progression from state to state. All processes in the continuous tumour growth model include tumour size as a continuous variable, which plays the major role in the modelling framework. The natural history submodels for tumour growth and symptomatic detection are adopted from previous works for data in absence of screening [87, 101, 102, 104]. The submodel for screening sensitivity comes from Weedon-Fekjær et al. [92, 106] whose work has also inspired the maximum likelihood estimation (MLE) procedure.

5.1 Modelling of three latent processes

Following is a description of the main modelling framework proposed in Study I, based on tumour size at diagnosis, without inclusion of covariates. The main model consists of five unknown parameters.

Submodel for tumour growth

The first submodel is for tumour growth, which is assumed to be exponential. To allow for heterogeneity in growth rates the model includes a gamma-distributed random effect for inverse tumour growth rate, R. Tumours are assumed to be spherical and t years after tumour onset, the volume (in mm^3) is specified as

$$V(t,r) = V_0 e^{t/r},\ t \geq 0, r > 0. \tag{5.1}$$

V_0 represents the tumour volume from when time starts to count (tumour onset) and may for instance represent the volume of one cell, or a small tumour volume not yet detectable by screening. r is a realisation from the random effects model with shape and rate parameters τ_1 and τ_2, such that

$$f_R(r) = \frac{\tau_2^{\tau_1}}{\Gamma(\tau_1)} r^{\tau_1 - 1} e^{-\tau_2 r},\ r \geq 0. \tag{5.2}$$

Submodel for time to symptomatic detection

The second submodel is a hazard function for time to symptomatic detection, assumed to be dependent on the tumour volume through

$$P(T_{det} \in [t, t+dt) | T_{det} > t, R = r) = \eta V(t,r)dt + o(dt),\ V(t,r) \geq V_0, \qquad (5.3)$$

where η is a constant and T_{det} is the time for symptomatic detection, in absence of screening.

Submodel for mammography screening sensitivity

The third submodel is for mammography screening sensitivity and enables modelling of data in presence of screening. It assumes the logistic form

$$S(d) = \frac{\exp(\beta_1 + \beta_2 d)}{1 + \exp(\beta_1 + \beta_2 d)}, \qquad (5.4)$$

in which β_1 and β_2 are constants and d is the tumour diameter in mm.

5.2 Natural history model covariates

As mentioned, the three latent processes described in Section 5.1 are all linked to tumour size. By using regular statistical techniques such as regression modelling, it is possible, in breast cancer cases, to find covariates that are associated with tumour size. Shedding light on the mechanisms behind the associations is however more difficult. Inclusion of covariates into the latent processes described herein, has the potential to increase knowledge regarding why some women are diagnosed with larger tumours than others. This makes our natural history model useful in the development of individual screening programmes. This Section presents the proposal (from Studies I and II) of how covariates may be included in the different parts of the continuous tumour growth model.

Submodel for tumour growth

For the gamma distributed inverse tumour growth rate (5.2), it holds that $E(R) = \frac{\tau_1}{\tau_2}$ and $Var(R) = \frac{\tau_1}{\tau_2^2}$. In Study II we propose to do a mean reparametrisation, so that $E(R) = \frac{\tau_1}{\tau_2} = \mu$ and $Var(R) = \frac{\tau_1}{\tau_2^2} = \sigma^2\mu^2$, in which σ is the constant coefficient of variation. We model

$$log(\mu) = \alpha_1 + \alpha_2 x_1 + ... + \alpha_{n-1} x_n, \qquad (5.5)$$

in which x_1 to x_n represents the covariates to be included. The parameters $\alpha_1, \alpha_2, ..., \alpha_{n-1}, \sigma$ will be estimated instead of τ_1 and τ_2. Using this parametrisation enables modelling of covariate effects on the mean inverse tumour growth rate. In Study

II, we include BMI as X_1, for unravelling whether the association between body and tumour sizes is due to tumour growth rate and/or symptomatic detectability.

Submodel for symptomatic detectability

For the hazard function (5.3) we propose in Study II, to allow for inclusion of covariates by letting

$$log(\eta) = \lambda_1 + \lambda_2 z_1 + ... + \lambda_{n-1} z_n. \tag{5.6}$$

Parameters to estimate are $\lambda_1, \lambda_2, ..., \lambda_{n-1}$. In Study II we include TBA (a determinant for breast size) as z_1.

Submodel for mammography screening sensitivity

As proposed in Study I, for the screening sensitivity function (5.4) covariates $q_1, q_2, ..., q_{n-2}$ may be included as

$$S(d, q_1, q_2, ..., q_{n-2}) = \frac{\exp(\beta_1 + \beta_2 d + \beta_3 q_1 ... + \beta_n q_{n-2})}{1 + \exp(\beta_1 + \beta_2 d + \beta_3 q_1 ... + \beta_n q_{n-2})}. \tag{5.7}$$

This is done in Study I (PD and interaction term with d), Study II (PD), Study III (PD and SI) and Study IV (PD), for quantification, testing or confounding purposes.

5.3 Estimation procedures

In Studies I-III, estimation of the unknown parameters in the three latent processes are based on maximum likelihood estimation, MLE. A likelihood function is the probability of the observed data, as a function of the unknown parameters [121]. Let $\boldsymbol{\theta}$ be a vector consisting of the parameters to be estimated and let $o_1, ..., o_n$ be realisations from the discrete random variables $O_1, ..., O_n$ following probability mass functions $P_1(O_1 = o_1; \boldsymbol{\theta}), ..., P_n(O_n = o_n; \boldsymbol{\theta}) = p_1(o_1; \boldsymbol{\theta}), ..., p_n(o_n; \boldsymbol{\theta})$. Then the likelihood function can be written as

$$L(\boldsymbol{\theta}) = \prod_j p_j(o_j; \boldsymbol{\theta}). \tag{5.8}$$

Further let $\hat{\boldsymbol{\theta}}_{ML}$ be the estimate of $\boldsymbol{\theta}$ that maximises the likelihood function. As the logarithm of a function is strictly monotone, an equivalent estimate can be found from maximising the log-likelihood function (often numerically more convenient) [122]

$$l(\boldsymbol{\theta}) = log(L(\boldsymbol{\theta})) = \sum_j log(p_j(o_j; \boldsymbol{\theta})). \tag{5.9}$$

The profile likelihood function divides the set of unknown parameters into parameters of interest $\boldsymbol{\theta}_i$ and nuisance parameters $\boldsymbol{\theta}_n$. In this function, the parameter values for the

nuisance parameters may be set to the values obtained from MLE of (5.8), so that these parameters are seen as fixed constants.

5.3.1 Point estimation

In this thesis $O_1, ..., O_n$ represent random variables, for the n included cases, of tumour size diameters (in mm) at diagnosis. The tumour size in each case j, follows an individual probability distribution P_j, conditional on detection mode, screening history and possibly other covariates. Further, tumour sizes are categorised into intervals I, indexed with i, and are assumed to follow multinomial distributions, so that

$$p_j(o_j;\boldsymbol{\theta}) \approx \prod_i P_j(O_j \in I_i;\boldsymbol{\theta})^{o_{i,j}}, \tag{5.10}$$

where $o_{i,j}$ is 1 if the tumour in woman j is detected in size interval i, and 0 otherwise. We let $P_j(O_j \in I_i;\boldsymbol{\theta})$ be written as $p_{i,j}$, which in words read: *the probability for the tumour in woman j to be detected in size interval i, given the detection mode and screening history of woman j*. Screening history is measured as the difference in days between the time of diagnosis and previous negative screenings. This implies that most of the probability distributions P_j are unique, except, for instance for symptomatic cases without screening history. The complete likelihood is written as

$$L(\boldsymbol{\theta}) = \prod_j \prod_i {p_{i,j}}^{o_{i,j}}, \tag{5.11}$$

and includes approximation and summation steps, thus no closed-form solution is available. A similar likelihood function setup was described by Weedon-Fekjær et al. [92, 106], with the main difference that, for interval/symptomatic cases, we propose modelling of tumour size, instead of modelling the incidence of interval cases (which requires data on all women at risk of getting breast cancer). For calculations of $p_{i,j}$, based on (5.1)-(5.4), see Section 5.4.1.

Optimisation procedure

Optimisation of (5.11) was carried out using the modified quasi-Newton optimisation procedure presented by Byrd et al. [123] by the *optim* function in the statistical program R [124], using the option *L-BFGS-B*. The method described by Newton for finding extreme values of functions, requires analytical solutions of e.g. the Hessian matrix, whereas the quasi-Newton methods, such as the BFGS-algorithm [125], use approximate values. The modified method used in this thesis has been developed to require less computational memory for optimisation of parameters.

5.3.2 Variance estimation and *p*-values

Due to a complex likelihood function, computer-intensive, numerical methods are used not only for point estimations, but also for evaluating the variability of the parameter estimates.

Likelihood ratio test

Let us assume that we have two nested models. We call the model with more parameters the alternative model and the model with fewer parameters the null model. Let L_1 be the optimised likelihood function value for the alternative model and L_0 the optimised likelihood function value for the null model. We define the likelihood ratio as

$$\Lambda = \frac{L_0}{L_1}. \tag{5.12}$$

Following an important result by Wilks [126] we then use that $-2\log(\Lambda)$ (for large n) approximately follows a χ^2_d-distribution, where d is the difference in dimensionalities of the models, to calculate p-values. We use this test to study the importance of parameter inclusion in Studies I-III.

Profile likelihood point wise confidence interval

In Studies I and II, the profile likelihood function was used to produce 95% point wise confidence intervals. The procedure is based on using the same (approximate) distributional assumption as for the likelihood ratio test above. When creating a point wise confidence interval for a parameter estimate $\hat{\theta}_{ML}$, the profile likelihood function was evaluated (optimised) for several different values of θ, to find for which values the MLE corresponded to a 95% confidence interval according to the quantiles of the χ^2_1-distribution [127].

Non-parametric bootstrap percentile confidence interval

Non-parametric bootstrapping is a resampling procedure, which may be used to estimate quantities such as the variance or confidence intervals of an estimator. Each bootstrap sample consists of n values, sampled with replacement, from the original sample of the same size n [128]. For each of the m drawn bootstrap samples the same estimation procedure (in this thesis MLE) is used as for the original sample, which creates a set of m bootstrap estimates for each parameter θ. When creating a 95% percentile confidence interval for θ, the range of the 95% least extreme values of the bootstrap estimates are presented, such that the same number of points are excluded from both sides of the interval. This method has been applied in Studies I-III.

Non-parametric bootstrap percentile test

In Study I non-parametric bootstrap percentile tests were carried out, using the same sampling technique as above. The p-value related to a null hypothesis such as $H_0 : \theta = 0$ was approximated by twice the value for the fraction of bootstrap estimates being below or above 0, according to the least common direction of the bootstrap estimates.

5.4 Calculations of conditional probabilities

The continuous tumour growth model framework presented in functions (5.1)-(5.4), along with additional assumptions of a *stable disease population*, make it possible to derive individual (conditional) distributions for different variables, both variables observable at diagnosis and latent, unknown variables. In comparison to unconditional distributions of variables that apply to all women in a population, conditional distributions may be helpful for studying individual screening programmes. In this thesis, some conditional probability distributions are presented. See Isheden and Humphreys [129] for further derivations. Observable variables may be used for derivation of likelihood functions (in this case for tumour size) and latent variables (such as lead time and tumour growth rate) may be used for bias corrections.

Assuming a stable disease population means that the rate of breast cancer onset is constant over time in the population and that the distributions for age at tumour onset and time to symptomatic detection is constant across calendar time.

5.4.1 Tumour size

The likelihood function (5.11), which includes the conditional probabilities $p_{i,j}$ for tumour size at diagnosis, resembles the likelihood function presented in Weedon-Fekjær et al. [92, 106]. They derived $p_{i,j}$ for screen-detected cases, but instead of modelling tumour size for symptomatic cases, they modelled the incidence of intervals cases. Calculations for tumour sizes at screen-detection were based on calculating backwards from a data set with tumour sizes in absence of screening (no model for symptomatic detection was assumed). In Study I, $p_{i,j}$ (also used in Study II) is presented for both screen-detected and symptomatically detected cases conditional on screening history. The derivations are based on back calculation from the symptomatic tumour size distribution in absence of screening (i.e. no external data set of tumour sizes in absence of screening is needed). In Study III a recently developed, less computer intensive derivation of $p_{i,j}$ [129] is used.

Studies I and II

Based on functions (5.1)-(5.3), Plevritis et al. [87] derived the marginal distribution for the volumes at symptomatic detection, $f_{V_{det}}(v)$, as well as the conditional distribution, given inverse tumour growth rate, $f_{V_{det}|R=r}(v)$. The latter is especially useful for simulating the natural history of breast cancer. In Study I, we present the reverse conditional distribution, using the lower incomplete gamma function, as

$$F_{R|V_{det}=v}(r) = \frac{\gamma(\tau_1+1, r(\tau_2+\eta(v-V_0)))}{\Gamma(\tau_1+1)},\ v \geq V_0, \tag{5.13}$$

which is applied in the back calculations from the symptomatic tumour size distribution in absence of screening to the tumour sizes at positive or negative screen exams, used in the derivations of $p_{i,j}$ (Section 5.3.1).

To briefly describe the conditional probabilities for tumour size ($p_{i,j}$), let woman j have a screening record of k screenings, then j_k is the time point for her last (positive or negative) screen exam, j_{k-1} is the time point for her second last screening, etc. and $\mathbf{B}^{\mathbf{c}}_{\mathbf{j_k}}$ is the event that a tumour in woman j was not detected at any of the screening rounds 1 to k. Further let $C_{i,j,t}$ be the event that a tumour in woman j at time point t is in size interval i (which corresponds to a tumour diameter of $d_{i,j}$). Also let $D_{j,t,f}$ imply that a tumour in woman j was supposed to be symptomatically detected f time units after the time point t in absence of screening, and $A_{g,j}$ be the event that the cancer would have become clinical for woman j in tumour size interval g, in absence of screening. (Using the index j in all events allows us to consider different covariate values for each individual, to be included in the three submodels – Section 5.2.)

Then for screen-detected cases it is shown in Study I that, approximately,

$$p_{i,j} \propto S(d_{i,j}) \cdot \sum_s P(\mathbf{B}^{\mathbf{c}}_{\mathbf{j_{k-1}}} | C_{i,j,j_k} \cap C_{s,j,j_{k-1}}) P(C_{s,j,j_{k-1}} | C_{i,j,j_k}) \cdot \sum_f \sum_g P(C_{i,j,j_k} | D_{j,j_k,f} \cap A_{g,j}) P(A_{g,j}), \tag{5.14}$$

where the sum over s (tumour size at last negative screening) is removed for screen-detected cases without previous screening exams. In the above formulae $S(d_{i,j})$ is the mammography screening sensitivity function (5.4) at the diagnostic screen exam. The sum over s includes the probabilities of masking at all previous screenings (conditional on tumour sizes at the screenings) and the probability that the tumour was in size interval s at the last screening before the diagnostic one. The double sum is an approximation of $P(C_{i,j,j_k})$, in which, the previously mentioned, back calculations from tumour size intervals at symptomatic detection, g, to tumour size intervals at the positive screen exam, i, are used, allowing for all lead times, f. For the double sum, distributions of V_{det} as well as $R|V_{det}$ (5.13) are used.

For symptomatically detected cases it is shown in Study I that

$$p_{i,j} \propto P(A_{i,j}) \cdot \sum_s P(\mathbf{B}^{\mathbf{c}}_{\mathbf{j_k}} | A_{i,j} \cap D_{j,j_k,f} \cap C_{s,j,j_k}) P(C_{s,j,j_k} | D_{j,j_k,f} \cap A_{i,j}), \tag{5.15}$$

where the sum over s is similar to the one in the conditional probability (5.14). Distributions of V_{det} and $R|V_{det}$ as well as the mammography screening sensitivity function (5.4) are used.

The derivation of $P(C_{s,j,j_{k-1}}|C_{i,j,j_k})$ in the conditional probability (5.14) is unfortunately not optimally presented in Study I, where it was calculated by using an approximation of the distribution $R|V_{det} > v$. An improved calculation of this probability could instead be

$$P(C_{s,j,j_{k-1}}|C_{i,j,j_k}) = \sum_r P(C_{s,j,j_{k-1}}|R \in r \cap C_{i,j,j_k}) P(R \in r|C_{i,j,j_k}), \tag{5.16}$$

where r represents an inverse tumour growth interval. For calculations of the latter probability, the derived distribution for $R|C = c$ shown in Isheden and Humphreys [129] may be used, where C represents the volume of a tumour in a non-screened, stable population at any point in time, i.e. not specifically at symptomatic detection such as for V_{det}.

Study III

In Study III a novel derivation of the conditional probability for screen-detected cases [129] is used, to increase computational efficiency, which was acquired by finding an analytical solution for the complex quantity $P(C_{i,j,j_k})$. This made it possible to remove the back calculation $\sum_f \sum_g P(C_{i,j,j_k}|D_{j,j_k,f} \cap A_{g,j}) P(A_{g,j})$ in (5.14). Also the derived probability $P(C_{s,j,j_{k-1}}|C_{i,j,j_k})$ by Isheden and Humphreys [129] was used, similar to the solution in (5.16).

5.4.2 Tumour growth rate and lead time

The conditional distribution for the latent variable inverse tumour growth rate, given tumour size at symptomatic detection (for women without previous screenings) (5.13) was derived in Study I. Isheden and Humphreys [129] derived the distribution for the inverse tumour growth rate, conditional on tumour size at screen detection for cases without screening history ($R|C$). In Study IV, based on their results, we show the distribution for the inverse tumour growth rate, conditional on tumour size at screen detection for cases *with* previous recorded screenings. A similar approach may be used for deriving the distribution for the inverse tumour growth rate, conditional on screening history and tumour size at *symptomatic* detection.

In Study IV we also present (individual) lead time (L) distributions for screen-detected cases, conditional on screening history and PD. Additional covariates, if jointly estimated in the continuous tumour growth model, are straightforward to include.

5.5 Lead time bias correction

Based on the conditional lead time distributions presented in Study IV we develop a procedure for correcting of lead time bias in survival comparisons. The procedure includes sampling a lead time, l^*, for each woman based on her individual distribution. If s represents the observed survival time measured from screen detection, then $s - l^*$ serves as an estimate of the survival time measured from when the symptomatic detection would have occurred in absence of screening. Conditional on observed survival times, the conditional lead time distributions also make it possible to estimate the probability for a woman to be overdiagnosed, which occurs when the mathematical relation $s - L < 0$ holds. The correction is adapted to handle women dying from breast cancer, but also censored women/women dying from other causes.

5.6 Simulation studies

Simulation studies are present in Studies I-IV to evaluate derived methodology. They have also been used to compare our methods to other described in the literature. In the simulations, life histories of breast cancer cases are produced, based on our described continuous tumour growth model, with a few exceptions. Focus has been on the time from tumour onset to detection by screening or symptoms, but, additionally in Study IV, breast cancer specific survival (in presence and absence of a screening programme) was simulated. Screen attendance is assumed complete throughout the studies and screening occurs every other year. The emphasis of the simulations has not been to make them fully realistic for evaluating screening programmes, but rather they are used as a tool to answer/show implications of specific methodological questions.

Dependence of tumour growth rate and symptomatic tumour size

In the studies by Weedon-Fekjær et al. [92, 106] a model for symptomatic detection was not assumed. This implies in our setting that distributions for V_{det} and $R|V_{det}$ cannot be derived. Instead, an external data set of tumour sizes in absence of screening may be used. Also, in the derivations of the likelihood function (i.e. in the probability $P(C_{i,j,j_k}|D_{j,j_k,f} \cap A_{g,j})$), use of the distribution of R instead of $R|V_{det}$, implies that the variables R and V_{det} are assumed to be uncorrelated. In Study I, under our continuous tumour growth model, we evaluate the implications of not assuming a submodel for symptomatic detection. For comparing our estimation procedure to a method without a symptomatic detection submodel, we compared MLE's using a distribution for $R|V_{det}$ instead of R and also using the derived distribution for V_{det} (which vary throughout the optimisation procedure as different sets of parameter values are evaluated) instead of a fix tumour size distribution based on simulated data.

Detection of image markers related to screening sensitivity

Strand et al. [76] used logistic regression with interval vs. screen-detected cancer as the depedendent variable (i.e. only screen attenders are included) to find image markers of MD from mammograms related to being an interval case. In a logistic regression, the log odds for being an interval case is modelled as

$$log\left(\frac{\pi}{1-\pi}\right) = \alpha_0 + \alpha_1 X_1 + ... + \alpha_p X_p, \tag{5.17}$$

where the variables $X_1, ..., X_p$ are the independent ones, $\pi = P(\text{interval case})$ and $1-\pi = P(\text{screening case})$.

In Study III we simulate data under our continuous tumour growth model, including an image marker affecting mammography screening sensitivity. We also simulate data under the tumour growth law (2.4) to take into account model misspecifications. We compare the statistical powers, to detect image markers related to mammography screening sensitivity, of the logistic regression model (with $p = 1$) and the approach of including a covariate in the mammography screening sensitivity submodel (5.4) of our continuous tumour growth model. The same set of women are included in both procedures. For generalisation purposes, several different inverse tumour growth rate distributions were used in the exponential tumour growth function (5.1).

Lead time bias correction on stratified data

Duffy et al. [96] proposed the use of a lead time bias correction for survival comparisons, based on the basic three-state Markov model (using exponentially distributed sojourn times), see Section 2.3.2. In the model, the lead time follows the same distribution as the sojourn time. Slightly simplified, the correction procedure is based on subtracting the mean, unconditional, sojourn time (which equals the mean lead time) from the screen-detected cases' survival times. For this reason we call this correction the *average* correction.

In Study IV we perform a simulation study based on our continuous tumour growth model, together with an exponential time to breast cancer death measured from the time the tumours would have been symptomatically detected in absence of screening. This is done to ensure that no women die before their lead times have passed. We stratify the screen-detected cases into equally sized groups based on the median tumour size at diagnosis in presence of screening. Then we perform the average correction by first estimating the mean sojourn time in the complete simulated data cohort in presence of screening [92], followed by subtracting it from each woman's survival time. On the stratified data we compare this procedure to the individual correction proposed in Study IV, and to the true values from the simulation. (A comparison on unstratified data was also performed).

Length-biased sampling and the effect of length bias on survival comparisons

In Study IV, to evaluate the length biased-sampling and to increase our understanding of the bias on survival comparisons, we perform a simulation study. To enable a length bias and an effect of the stage shift, we make breast cancer specific survival dependent on the tumour size at diagnosis and the tumour growth rate. Survival times are recorded for all women under two different counterfactual scenarios; in the presence, S_{pres}, and absence, S_{abs}, of a screening programme. As sojourn times are not defined in continuous tumour growth models, we instead propose the use of *tumour presence times* which is the time from when the tumour was of a specific small tumour size (we use a diameter of 2 mm) until the time it would have become symptomatically detected, in absence of screening. Based on our continuous tumour growth model and parameter estimates presented in Isheden and Humphreys [129], we compare tumour presence times between screen-detected and interval cases to evaluate the length-biased sampling. We also compare the factors affecting it; in our model inverse tumour growth rates and symptomatic tumour sizes.

To study the effect of length bias on survival comparisons, together with the effects of stage shift and lead time, we compare different measures of survival times. Specifically we compare observed survival for screen-detected cases, lead time corrected survival for screen-detected cases, survival for screen-detected cases in absence of screening and observed survival for interval cases.

5.7 Computational aspects and practicalities

Programming for all studies in this thesis has been done by using the statistical program R [124]. All of the studies use methods that are computationally intensive, both for producing simulated data sets, but most of all in the evaluations of the likelihood function (5.11). This implies that applying bootstrap methods for producing variance estimations, takes a long time. Analyses in Studies I-III have been especially time-consuming. Inclusion of covariates in the inverse tumour growth rate distribution and time to symptomatic detection function involve additional computations compared to including covariates in the mammography screening sensitivity function (Study II). A large number of likelihood evaluations were required in the simulation study (Study III).

However, computational speed was much higher (computational time being reduced by 88% [129]) in Study III, than it was in Studies I and II, due to the fact that, in Study III, back calculation in the likelihood evaluation procedure could be removed by the inclusion of a closed-form solution.

6 Main results

In this Section the main results from Studies I-IV are presented; firstly the methodological evaluations, followed by quantifications and tests based on observational data and lastly simulation-based results.

6.1 Methodological evaluations

6.1.1 Joint estimation of three latent processes

In Study I we found that when simulating data in presence of screening, under our proposed continuous tumour growth model, we could learn the correct values of the parameters within the submodels of tumour growth rate, screening sensitivity and time to symptomatic detection. In Study II we evaluate the effect of inclusion of covariates in different parts of the model, especially the inverse tumour growth rate and the time to symptomatic detection submodels. Data was simulated under our continuous tumour growth model, including a covariate in either, or both, of the two submodels. Effects were identifiable as long as the covariate was included in both submodels.

As described in Section 5.6, by not assuming a submodel for time to symptomatic detection in the continuous tumour growth model framework described by Weedon-Fekjær et al. [92, 106], the likelihood function evaluations have to be based on the assumption that tumour growth rate and symptomatic tumour size are independent variables. Under our proposed continuous tumour growth model, we could show that, basing the likelihood functions on such an assumption leads to biased estimates of tumour growth rate and screening sensitivity. Our model relaxes this assumption.

6.1.2 Lead time and bias correction

In Study IV we show the derived individual lead time distributions, conditional on tumour size at screen-detection, screening history and PD. Other covariates are straightforward to include/condition on. To illustrate this, individual lead time distributions based on TBA and tumour sizes for screen-detected cases with no screening history are shown in Figure 6.1, based on parameter estimates from Study II (allowing for TBA to affect time to symptomatic detection and BMI to affect inverse tumour growth rate). TBA is standardised based on the observational data used in Study II.

Further, in Study IV, we make use of simulated data under our continuous tumour growth model, together with exponentially distributed times to breast cancer death (measured from the time of when symptomatic detection would have occurred in absence of screening). We show that the proposed lead time correction (Section 5.4.2)

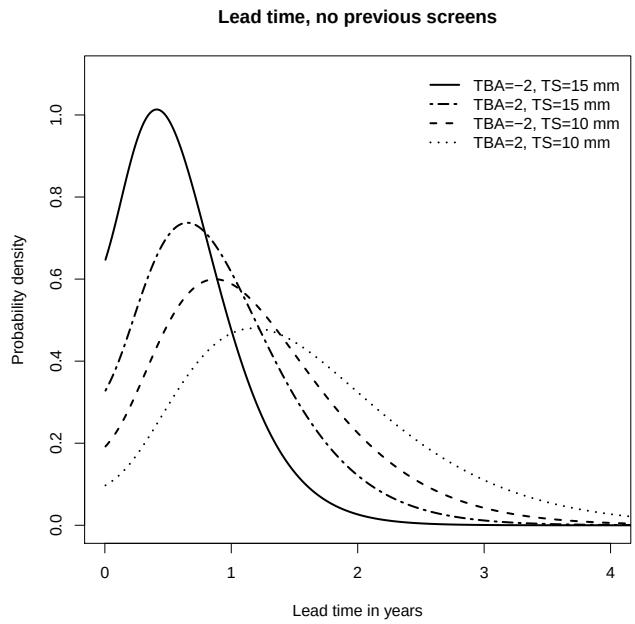


Figure 6.1: Individual lead time distributions for screen-detected cases based on tumour sizes (TS) of 10 and 15 mms and standardised total breast areas (TBA) of -2 and 2, based on the inverse tumour growth rate distribution according to the mean BMI seen in Study II.

gives unbiased survival times measured from symptomatic diagnosis, for all screen-detected cases grouped together, as well as stratified on tumour size at diagnosis (being below or above the median value).

For the *average* correction (see Section 5.6) using unconditional lead time distributions, lead time corrected survival times for screen-detected cases, stratified on being below or above the median tumour size at diagnosis, were biased. In general, use of unconditional lead time distributions results in an undercorrection for small tumours (taking away too short lead times) and an overcorrection for large tumours (taking away too long lead times). This result can also be extended to other variables than tumour size, that both affect the lengths of lead time and survival time.

6.2 Quantifications and tests

For the overall continuous tumour growth model, the most updated parameter estimates for tumour growth rate, mammography screening sensitivity and time to symptomatic detection, are presented in Study III. In Figure 6.2 these are summarised (using the median value of SI in the population for plotting screening sensitivity). The quantification of tumour growth (to the left) is presented as the time in years it takes a tumour to increase in diameter, relative to the time it was 15 mm in diameter. It can be seen that the individual-to-individual variation in growth rates is estimated to be large. On the right-hand side of Figure 6.2 probabilities of detecting a tumour are shown, comparing screening sensitivity to symptomatic sensitivity. Screening sensitivity refers to the probability of tumour detection at a screen exam (not a process in continuous time). The symptomatic sensitivity is presented as a cumulative probability that the tumour would have been detected by symptoms at a specific tumour size, in absence of screening (conditional on the estimated median tumour growth rate).

6.2.1 Mammography screening sensitivity in light of mammographic density

Mammography screening sensitivity has been quantified in Studies I-III. In Study I we presented novel estimates of the relationship between PD and tumour size (considered jointly) with mammography screening sensitivity. A high PD was found to lower screening sensitivity (p-value=$4 \cdot 10^{-3}$). When both PD and tumour size had been included in the model, an interaction effect between the two was not significant. In Study II, controlling for body size in the submodels for tumour growth and time to symptomatic detection, had the effect of narrowing the confidence intervals for all parameters included in the sensitivity function. PD was significant (p-value=$6.3 \cdot 10^{-4}$). In Study III, without controlling for body size, the corresponding p-value for PD was $5.7 \cdot 10^{-3}$. The role of PD and tumour size in mammography screening sensitivity is

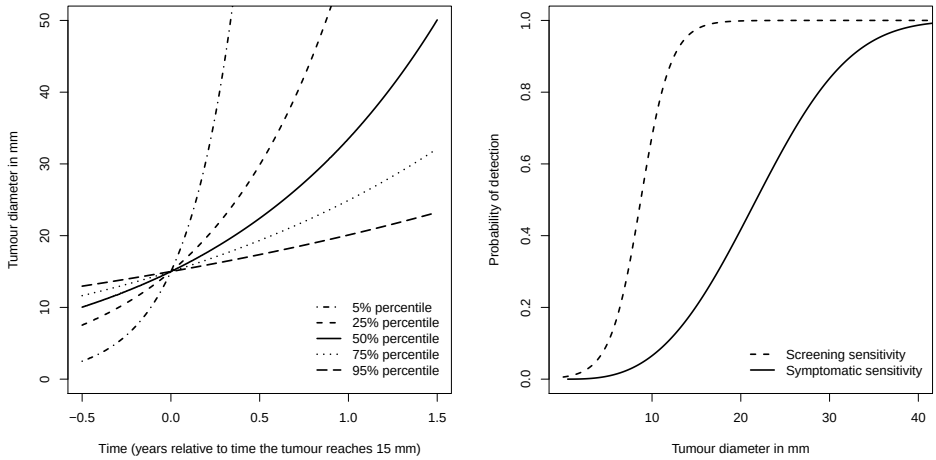


Figure 6.2: Quantification of tumour growth heterogeneity (left) and screening and symptomatic sensitivites (right).

presented in Figure 6.3 (to the right). 95% confidence intervals, received from 200 nonparametric bootstrap samples are also presented. The PD values of 0 and 100% have been plotted for comparison purposes, although in the data set, no woman has a value higher than around 85%.

In Study III a similar quantification, but for the role of SI (measuring scatteredness of mammographically dense tissues) in mammography screening sensitivity was presented (see Figure 6.3 to the left). Since SI was normalised, it ranged between 0-1, but higher values were more common than lower (see Study III). Comparing the quantifications for PD and SI in mammography screening sensitivity, it can be seen that the use of SI better distinguishes between which women that have a high compared to low sensitivity at different tumour sizes. P-values from likelihood ratio tests, explaining the additional effect of an image marker, when the other one was already included in the sensitivity model, were 0.76 for PD and $5.6 \cdot 10^{-4}$ for SI.

6.2.2 The association between body size and tumour size

Studies based on regular statistical regression models have shown positive associations between body size and tumour size at diagnosis [32, 34, 48]. However, with these standard techniques it is not possible to draw any conclusion regarding the reason(s) behind the associations. In Study II we include BMI as a covariate in the model for inverse tumour growth rate and TBA as a covariate in the model for time to symptomatic

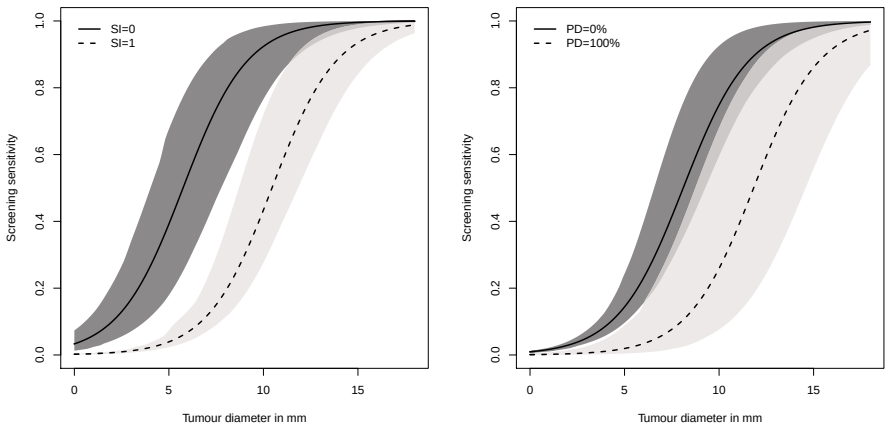


Figure 6.3: Quantification of mammography screening sensitivity as a function of tumour size and normalised SI compared to PD (without controlling for body size in the other submodels), plotted with 95% confidence intervals.

detection. We also include PD in the screening sensitivity function as a confounder, since PD is negatively correlated to body size. The estimates point in the directions that a high body size both delays the symptomatic detection and makes tumours grow faster. The p-value for the joint association test (H_0 : body size not associated with either process, H_A : body size associated with at least one of the processes) was $5.0 \cdot 10^{-5}$. Individually, only the association with symptomatic detectability reached statistical significance (p-value=0.022 for symptomatic detection and p-value=0.089 for inverse tumour growth rate).

6.3 Simulation-based studies

6.3.1 Detection of image markers related to screening sensitivity

In Study III, we compared the use of univariate logistic regression with interval vs. screen-detected cancer as the dependent variable, to our continuous tumour growth model setting (including a covariate in the screening sensitivity function), for detecting image markers related to the masking of tumours (Section 5.6). Comparisons were made on simulated data, which included a hypothetical image marker that affected screening sensitivity. Results showed that the statistical power to detect markers related to screening sensitivity was much larger for the continuous tumour growth model than it was for the logistic regression model. This was true when simulating data under our

continuous tumour growth model, using varying underlying speeds of growth, as well as when simulating under a logistic tumour growth law.

6.3.2 Length-biased sampling and the effect of length bias on survival comparisons

In Study IV we simulate data to study length-biased sampling and the effect of the bias on survival comparisons between e.g. screening and interval cases. Under our proposed continuous tumour growth model screening cases had longer tumour presence times (see Section 5.6) than interval cases. This was (partly) caused by interval cases and screening cases having different inverse tumour growth rate distributions – interval cases had on average faster growth rates than screening cases. However, we could also demonstrate another cause for the difference in tumour presence time distributions (which is seldom discussed in the literature) – that interval and screening cases had different distributions for symptomatic tumour size (the size at which a tumour would have been detected, in absence of screening). In general, for a tumour to be screen-detected, it is required that the tumour is not detectable by symptoms too early in the disease progression. As both the symptomatic tumour size and the inverse tumour growth rate affect breast cancer specific survival, both of these factors are a part of the length bias. Different conclusions can be drawn from this. One being that it may not be correct to assume that the difference between interval and screen-detected cases in tumour size distributions (in the presence of screening) is a result of the stage shift. However, it is likely that a large proportion of the difference is a cause of the earlier detection.

Regarding the effect of the length biased-sampling on survival comparisons between interval and screen-detected cases, novel theoretical conclusions can be drawn based on our model. Although it is likely that the length bias gives a survival advantage for screen-detected cases in comparison to interval cases, there is a hypothetical chance that the opposite is true. This can occur if the tumour growth rate does not affect survival to any large extent, whilst tumour size at diagnosis does, and that there is a group of women with a delayed symptomatic onset. In addition, a length bias correction procedure, to be used on survival comparisons, should not only be based on tumour growth rate, but should preferably also take symptomatic tumour size into account.

This is a simplified scenario; it is likely that the length bias is affected by other factors as well, such as the complete tumour stage and biology.

7 Discussion

Many of the results presented in this thesis are based on the assumptions of the same underlying natural history model, including an exponential tumour growth with a gamma distributed inverse tumour growth rate, a tumour volume-dependent hazard function for time to symptomatic detection and a logistic function for mammography screening sensitivity dependent on the tumour diameter. As is the case for all model assumptions used for explaining complex real-life processes/events, the ones used in this thesis will not be exact, but hopefully are approximate enough to be useful. It however seems reasonable that results from different statistical models together provide stronger, more robust evidence, than what can be provided from one model alone, as has been the line of thought adopted by the CISNET consortium [63, 65].

The described natural history model was partly chosen for mathematical tractability – with its use many processes may be derived and viewed as analytical expressions, on an individual (conditional) level. In addition, analytical approaches for testing for associations of covariates in all parts of the model can be developed. It is likely that a simulation-based approach for calibration would have to be developed instead of an analytical estimation procedure, if e.g. the exponential growth law, should be changed to some more mathematically complex growth law. The reason being that many of the analytical expressions used for estimation would not be possible to derive; for example the distribution for tumour sizes at symptomatic detection in absence of screening.

Our model is not as complete in terms of the number of processes included/estimated, as the models used in the CISNET consortium [63, 65]. It is however unique and has the potential to be developed further. All of the CISNET models except one, are based on simulation-based approaches for calibration. The one analytical approach [85] is based on a (partly Markovian) multi-state model rather than a continuous tumour growth model. However, our model does not differ in terms of the exponential tumour growth law assumption, comparing it to the three models assuming a continuous tumour growth. Regarding the CISNET models, an interesting note is that the MISCAN-Fadia model described by Tan et al. [110] is based on a continuous exponential tumour growth model, but was originally developed as the MISCAN model, a discrete state progression model presented in for example Habbema et al. [130]. The transition in models was made to improve the description of biological mechanisms, by the inclusion of a continuous tumour size. It was also made to enable a separation of processes of tumour growth, and clinical and screen detection.

Natural history model estimates differ in published studies, depending on both the data utilised, but also the statistical modelling and estimation techniques used. Not all of the quantifications in this thesis can be compared to other estimates. This is especially true for the quantifications of BMI on growth rate (Study II), TBA on symptomatic detection

(Study II) and SI on mammography screening sensitivity (Study III), as these are to the best of my knowledge, the first of their kind. However, BMI was estimated to be an inverse promoter of clinical disease, using a multi-state Markov model setting, in a study by Yen et al. [83]. As previously mentioned, the reason for a variable to be a promoter can differ. A variable may be an inverse promoter if it has either a slow growth rate, or a delayed symptomatic detection, or an earlier possibility of screen-detection. For body size, the two latter processes have been hypothesised (due to the correlation with PD). This result is however in line with our finding that there is a higher evidence for a delayed symptomatic detection in comparison with a faster tumour growth for women with higher BMI. However, in general, as multi-state models do not separate time into growth rate and symptomatic detectability, they are not well suited for explaining reasons behind associations between covariates and tumour size.

Our estimates of the role of tumour size in mammography screening sensitivity differ to some extent from the ones presented by Tan et al. [89]. They used a 13-state Markov model with tumour size categories (diameter in mm) of ≤ 10, $11-20$, $21-50$ and ≥ 51 and estimated corresponding screening sensitivities of 90, 91, 92 and 93% (in DCIS tumours the sensitivity was estimated to be 88%). Our results suggest that tumour size plays a more important role for screening sensitivity. It is likely that estimates vary due to the different statistical models used, but also due to the large tumour size intervals used by Tan et al [89]. It would have been interesting to see their quantification of sensitivity for tumours with a diameter of ≤ 5 mm. In comparison to the studies by Weedon-Fekjær et al. [92, 106], which assumed the same parametrical model for screening sensitivity, our estimate lies in between their two published ones.

There are fewer comparisons to make regarding our estimate of the role of PD in mammography screening sensitivity, especially when also considering tumour size. In 2004, Berg et al. [70] quantified screening sensitivity for the BI-RADS categories, but without taking tumour size into account. They estimated a 60% sensitivity in extremely dense breasts and a 100% sensitivity in predominantly fatty breasts. However, their sample size was small (110 breast cancer cases). Based on the three-state basic Markov model, Chiu et al. [131] reported a sensitivity of 62.8% in dense breasts (Tabár patterns IV and V) and a sensitivity of 82.0% in non-dense breasts (Tabár patterns I-III). Estimation of screening sensitivity and mean sojourn time was however not performed jointly. More common in multi-state Markov models is to estimate age-specific sensitivities [85, 132], which are not directly comparable to our estimates.

According to the quantification of tumour growth presented in Section 6.2, the median doubling time in days is estimated to be 105 days. Weedon-Fekjær et al. [92, 106] presented median doubling times of 99 and 152 days for tumours at 15 mm, assuming a logistic growth law. This growth law however leads to shorter doubling times for smaller tumours and longer doubling times for larger tumours. Corresponding estimates of the mean sojourn time from multi-state Markov models also tend to vary between studies

and data sets [90, 133, 134]. These are, as discussed, not directly comparable to our estimates of growth rate. However, they may be used as reference values for mean lead times (lead times follow the same distribution as sojourn times in memoryless Markov models). Our estimate of the length of lead times is shorter than many estimates of the mean sojourn times, but in line with the shorter quantifications presented. Our lead time estimate also resembles the estimate of mean lead time being close to one year, presented by Chen et al. [135] (but not, for example, the four years adopted by Duffy et al. [96] in their average lead time correction procedure).

Using a similar continuous tumour growth model, but for data in absence of screening and with the constraint that the expected inverse tumour growth rate is equal to one, Plevritis et al. [87] estimated that symptomatic detection occurred later in the disease progression than what is suggested by our estimates. Our model claims that tumours growing with a median growth rate are not usually symptomatically detected after a tumour size of 40 mm, in comparison to 60 mm, based on their parameter estimates. The difference in estimates may be caused by both differences in model assumptions (fixing the mean inverse tumour growth rate to one, or not) as well as non-exchangeable data sets. Breast cancer cases were diagnosed between 1975 and 1981 in the study presented by Plevritis et al. [87] whereas women in CAHRES were diagnosed between 1993 and 1995.

One of the potentials with the proposed continuous tumour growth model is its ability to shed further light on traditional concepts, such as the length bias. In Figure 7.1 the length biased-sampling is depicted as being caused by the rate of tumour growth and symptomatic tumour size, to be compared with Figure 2.10 in which only the dimension of (sojourn) time has been plotted. In Figure 7.1 the screen-detected case has a longer tumour presence time than the two depicted interval cases, although the reasons behind this differ; one of the interval cases grows faster but has the same symptomatic tumour size as the screen-detected case, the other interval case is easy-detectable but grows with the same speed as the screen-detectable case.

For practical reasons (including computational times) it was unfortunately not possible to present updated parameter estimates using the probability (5.16) in the conditional probability for tumour size (5.14), for results presented in Studies I and II. However, the parameter estimates of Study III are similar to the published estimates in Study I. It seems reasonable that the estimates in Study II would not change much either. In any future study of the role of body size for growth rate and symptomatic detectability, it would be interesting to also include, for instance, HRT as a confounder. This would now be practically possible in our continuous tumour growth model setting as the computational times have been improved upon [129].

Breast cancer is a heterogeneous disease. The natural history model presented herein is simplistic, but can be used to form the basis of a novel statistical model. To be of

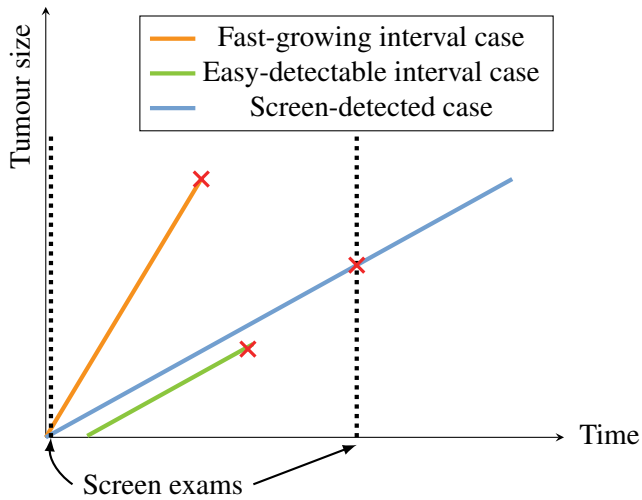


Figure 7.1: An illustration of length biased-sampling in two dimensions, where the time the tumour is present in a woman's body depends on the tumour growth rate and symptomatic detectability. The crosses represent times for detection. The screen-detected case has a slow growth rate and a delayed symptomatic detection.

higher relevance for future research it should be developed further. There are however unlimited expansions possible for the model in terms of including a complete tumour stage (with a possibility for tumours to stagnate in progression), a submodel for tumour onset and also one for breast cancer specific survival. Analytical expressions for all these processes are however challenging to derive, but once done, use of covariates within these processes may provide a comprehensive picture of the natural history of breast cancer. Such a model can be used for different purposes, for example to evaluate benefits and harms from screening programmes, either by using simulation studies or by estimating effects directly from observational data. The model can also be used for jointly testing for associations between risk factors/covariates in different latent processes.

8 Concluding remarks

In this thesis the basis of a novel natural history model for data on breast cancer cases in presence of mammography screening, has been described. The model is called a continuous tumour growth model and makes it possible to jointly estimate latent continuous processes of breast cancer tumour growth, time to symptomatic detection and mammography screening sensitivity. All parts of the model have the possibility of including covariates. In comparison to multi-state models, it has the advantage that it divides sojourn time/transition rates from preclinical to clinical disease, into processes of tumour growth rate and time to symptomatic detection. This enables further understanding of the natural history of breast cancer and e.g. increases knowledge of reasons behind covariates being promoters of clinical disease. Also, the model is superior in statistical power for detecting image markers related to masking of tumours, compared to univariate logistic regression with detection mode (interval vs. screen-detection) as the dependent variable.

The model framework makes it possible to derive individual (conditional) distributions of latent variables, rather than distributions on a population level (unconditional), an example being the lead time. The conditional lead time distributions derived in this thesis suggest that especially tumour size at screen diagnosis is of great importance for predicting when symptomatic detection would have occurred in absence of screening. We have shown how the conditional lead time distributions derived in this thesis can be used for correction of the lead time bias in survival comparisons between for example screen-detected and symptomatically detected cases. The conditional distributions are especially useful in stratified analyses. For the length bias, according to our continuous tumour growth model, the time a tumour is present in a woman's body is dependent on the tumour growth rate, but also the symptomatic tumour size. The latter is not usually discussed in the literature describing this bias. Since tumour growth rate and tumour size have their own relationships with survival, a length bias correction including both parts is superior to one being based on sojourn time. The extended view of this bias is also of importance for some interpretations of analyses affected by the length bias.

In applied work on postmenopausal breast cancer cases, we could quantify the relationship between tumour size and the probability that a tumour will be detected at screening. We could also show that percentage mammographic density, a quantitative image marker of mammographic density, significantly improves predictions of sensitivity. The image marker however does not take into account how the dense tissues are distributed in the breast. A measure of the scatteredness of dense tissues, referred to here as skewness of the intensity gradient, may have a higher ability than percentage mammographic density to distinguish between which women that have high or low mammography screening sensitivity.

Women with larger body size have an increased mammography screening sensitivity (due to having a lower percentage mammographic density), still their tumour sizes at diagnosis are larger. We could conclude that this is (partly) caused by women with larger breasts having difficulties in symptomatically detecting their tumours, or alternatively, delaying their visits to health care. Although not statistically significant, our results point in the direction that this is also caused by a faster tumour growth rate for women with higher BMI.

9 Future perspectives

Nowadays breast cancer screening programmes (with some exceptions) invite all women within a specific age range to be screened with the same frequency. It is becoming clear, however, that this may not be the best use of health care resources. Researchers are discussing how screening programmes could be implemented on a more individualised level [136]. Overall (long-term) breast cancer risk for a woman is probably the major factor that could be useful in deciding which women that should be screened and how often, but other factors (risk of particular subtypes), body size, may also be important. Implementation of individualised programmes could indeed improve the balance of benefits and harms from screening that women encounter, but pragmatically could be difficult or costly to implement.

Studying the processes discussed in this thesis, of screening sensitivity, symptomatic detectability and tumour growth, may shed light on issues around individualising screening programmes even further. The following is a simplified discussion of these processes and how knowledge about them could be useful for thinking about individualised screening.

For a screening programme to be efficient, and balance benefits and harms, women with a high enough risk of dying from breast cancer should clearly be screened. For each woman attaining such risk level, screening with a tool having a high sensitivity needs to be offered. As use of mammography screening is the golden standard in today's health care, individual quantifications of mammography screening sensitivity may be able to point out (probably together with analyses of cost-effectiveness) which women that should be offered another type of screening modality to increase sensitivity. Use of image analysis and rich data sources including mammograms, can help in further distinguishing screening sensitivities between (and within over time) individuals, as exemplified in Study III using the scatteredness of mammographically dense tissues to explain screening sensitivity, rather than the fraction of mammographically dense tissues in the breasts.

To some extent, for women not expected to die from other causes any time soon, the risk of dying from breast cancer in absence of screening, can be explained by processes of tumour growth rate and symptomatic detectability. However, other factors of tumour heterogeneity that are linked to the choice of treatment, such as molecular subtypes, are of great importance. These will not be included in this short and simplified discussion, as they have not, as of yet, been studied in this continuous tumour growth model. Other factors considered, fast growing tumours may, on average, have worse prognosis than slow-growing tumours. Fast growing tumours also have shorter windows of opportunity for being screen-detected. Individual quantifications of risk of getting fast-growing breast cancer, may help in distinguishing which women should be screened with short intervals.

Even then, screening may not be well-suited to fast-growing cancers. Therefore, if the risk of such cancer is high, preventive actions may also be of importance. In women with risk of only slow-growing breast cancer, a less frequent screening programme may be useful. If in addition such a woman has a high chance of early symptom onset (i.e. of finding the tumour by symptoms whilst it is still small) she may not need to participate in screening. For example, this may be the case for women with very small breasts. For all women, having a risk of delayed symptomatic onset *and* of getting fast or slow-growing breast cancer, screening is probably a useful tool.

In a recent study by Cohen et al. [137] a novel blood test called CancerSEEK, which can be used for screening eight different cancers, was proposed. The blood test was taken on cases having different types of cancer. For some cancer types, the sensitivity was high, but for breast cancer it was only 33%. However, in a future individual screening programme of breast cancer, a similar but improved blood test could potentially be of use as a complement to other screening modalities.

10 Acknowledgements

This journey at Karolinska Institutet started with my Master thesis in 2011, and seven years later, I have now finished my Doctoral thesis. Doing PhD studies at the same time in life as starting a family has truly been challenging, but with the help and support from many of you, whom I now would like to thank, it was possible.

I have truly had a great main supervisor. Thank you **Keith Humphreys** for introducing me to this topic, and for letting me do both a Master and a Doctoral thesis with you. You have given me much of your time, patience and guidance throughout these years and you also made it possible for me to complete my studies although the time schedule has been changing. Thank you for sharing research discussions, improving my writing skills and for always pointing me in the right direction. This thesis would not exist without all the help you have given me.

Going from a world of mathematical statistics only, to a hybrid one including also medicine, has not always been easy. *Keith, can you tell me once more what she just said but in human language?* my supervisor **Kamila Czene** said at one of our meetings. Thank you Kamila, for helping me to improve my ability to discuss research. Thank you also for sharing your, always so insightful thoughts, on my research. You have an ability to find the most intriguing research questions.

To **Per Hall**, the most famous one of my supervisors. All your efforts in producing high-quality data (a dream for statisticians to use) and your interest in explaining research to the general audience are truly inspirational. Thank you for that and for always being fast on answering questions.

My very intelligent supervisor **Mark Clements**, I am truly inspired by your programming skills and at some point in life I will probably work on my own skills as well. Thank you for always having a smiling face, for always being ready to help and for creating a good atmosphere in the Biostatistics group.

I am thankful for having been in a creative and inspiring research group. To the former group members **Hatef Darabi**, **Abbas Cheddad** and **Therese Andersson**, thank you for your kindness and for doing inspirational research. Hatef, thank you for giving me a kind welcoming to MEB. To the present group: I have truly appreciated our group meetings and chats in the corridor. Thank you **Gabriel Isheden** for sharing your positive view on life and for your insightful thoughts on research. I remember the time when Keith and I were joking about being tired from much work and you started a high-level theoretical discussion with the words: *Let us do a little thought experiment*. Thank you **Maya Alsheh Ali** for always smiling and being friendly, for teaching me some image analysis and for

helping me with the mammograms included in this kappa. Thank you also to our newest member **Rickard Strandberg** for your humility and for sharing interesting thoughts on research. I am also amazed by your ability to feel the taste of ramson.

To my dearest friends over the years at the department, **Alessandra Grotta** and **Anna Plym**. What would I have done without you? Thank you both for nice conversations and for always being friendly. It has been a true joy spending time with you.

To my past and present office mates. I would like to especially thank **Wengjiang Deng** for always being kind, having fika ready and for reminding me to have lunch, **Xingrong Liu**, for sharing your thesis templates and for always smiling and being encouraging even at times when you had much to do, **Henrik Olsson**, for always taking your time to talk. Thanks also to **Robert Szulkin**, **Johan Zetterqvist**, **Fredrik Jonsson** and **Louise Eriksson-Bergman**.

Thank you all who have been involved in the table tennis competitions at MEB. Especially I would like to thank **Marie Reilly** for helping me to organise the ladder once and for being a lovely person, and **Johan Rosén** for many great battles, for being helpful and friendly.

It has been a true joy being a part of the Biostatistics group at MEB. I would like to thank all of you, for creating such a good atmosphere in the group and for sharing much of your knowledge, including, previously not mentioned, but not limited to **Marie Jansson** (for helping out with many administrative tasks), **Elisabeth Dahlqwist**, **Daniela Mariosa**, **Caroline Weibull** (for introducing me to Keith), **Myejongjee Lee**, **Maria Anna di Lucca**, **Hannah Bower**, **Anna Johansson**, **Cecilia Lundholm**, **Annika Tillander**, **Sara Ekberg**, **Paul Dickman**, **Juni Palmgren**, **Paul Lambert**, **Yudi Pawitan**, **Andreas Karlsson**, **Thorgerdur Palsdottir**, **Martin Eklund**, **Setia Pramana**, **Alex Ploner**, **Sven Sandin**, **Rino Bellocco**, **Suo Chen**, **Teo Shu Mei**, **Shuang Hao**, **Aminata Ndiaye**, **Rose Bosire**, **Henrik Winell**, **Robert Karlsson**, **Sandra Eloranta**, **Peter Ström**, **Arvid Sjölander**, **Sophie Debonneville**, **Julien Bryois**, **Zheng Ning**, **Giorgio Tettamenti**, **Bénédicte Delcoigne**, **Mikael Franko**, **Erin Gabriel**, **Shengxin Liu**, **Agnieszka Szwajda**, **Nghia Vu** and **Li Yin**.

To the PhD student group at MEB, thank you for sharing this journey with me. I would especially like to thank **Fei Yang**, **Johanna Holm** and **Vilhelmina Ullemar** for being great friends, **Fredrik Strand** and **Jiangrong Wang** (also Xingrong, Hannah, Anna and Cecilia) for helping out and giving advice regarding the dissertation procedure. Thank you Andreas, Zheng and **Shadi Azam** for being on my pre-dissertation committee. I am also thankful to all other friendly, inspiring persons at MEB, too many to mention. Thank you especially to **Amelie Plymoth**, **Fredrik Wiklund**, **Henrik Larsson**, **Camilla Ahlqvist**, **Gunilla Nilsson-Roos**, **Lina Werner** and all **TA staff** for always helping out.

To the teachers at Uppsala University who inspired me to continue my journey in mathematical statistics. Thank you **Silvelyn Zwanzig** for your nice personality, all the great courses you held and for being my Bachelor thesis supervisor. To other great teachers, thank you **Allan Gut**, **Sven-Erick Alm**, **Måns Thulin** and **Jesper Rydén**.

Tack till **Marina**, **Bruno**, **Karin**, **Ove**, **Elina** och **Fredrik** för att ni är härliga, omtänksamma vänner som jag uppskattar att umgås med. Tack även till mina nära och kära, bland andra **Conny**, **Emanuel**, **Daniella**, **Johnny**, **Birgitta** och **Camilla** med familjer, för släktkalas, hjälp och umgänge. Ett speciellt tack till mina svärföräldrar **Peter** och **Britt-Christine** för all hjälp och för att ni är underbara farföräldrar till barnen, det är alltid härligt att träffa er. Tack till **Mormor** och **Farmor**, för att ni är starka kvinnor som ger mig perspektiv på livet.

Att ha två systrar är en av de största gåvorna jag fått i livet. Tack **Sandra** för att du alltid finns där för mig, redo att lösa alla världens problem eller bara umgås, vi har följts åt genom hela livet. **Moa**, du är guld värd. Tack för all den glädje du ger mig och dina fem syskonbarn och för alla våra innerliga, givande samtal. Tack även till **Andreas** för hjälp med barnen, för att du alltid är redo att bolla nya idéer och för din gästvänlighet. Tack **Arvid**, **John** och **Vera** för att ni är så fina syskonbarn.

Tack till mina föräldrar för all kärlek och inspiration jag fått i livet, och för all hjälp ni ger min familj. Tack **Pappa** för att du förberett mig inför den här resan genom alla idrotter vi tränat tillsammans, särskilt bordtennis. Det finns många likheter med att förbereda sig till stora tävlingar som till en disputation; målbild, förberedelser och envishet. Till **Mamma**, min klippa, tack för all den tid du spenderat med oss barn när vi var små och för att du den här sommaren har passat Aron och Märta många gånger när jag behövde skriva på avhandlingen.

Till min man **Sebastian**. Tack för att du finns där för mig och inte stoppar mina galna, kreativa idéer (mer än nödvändigt). Tack för att du åt en chili men sa nej till att flytta till Grönland. Tack för din kärlek och för att du har stöttat mig att fullfölja mina studier. Nu gör vi oss redo för nya äventyr tillsammans.

Till de viktigaste jag har. Tänk att barn ibland kan förstå så mycket mer än vuxna. Att få vara mamma var länge min högsta dröm, och jag kan inte tänka mig något bättre än att få vara *er* mamma. Tack **Aron** för att du är en sådan inkännande, hjälpsam storebror (*Mamma, jag låter Märta välja barnprogram för att hon är ledsen*), för att du delar med dig av dina intressanta funderingar om livet och för att du vågar prova nya saker med mig. Tack **Märta** för att du är en sådan livlig glädjespruta, för din förmåga att visa kärlek och uppskattning, och för dina spännande kommentarer. Som till exempel när du fick veta att morfar skickar sina tjurkalvar vidare till en annan bonde, då sa du efter en stund: *Det är godare gräs hos den andra bonden*.

References

[1] Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. *Lyon, France: International Agency for Research on Cancer* 2013; available from: http://globocan.iarc.fr, accessed on 05/07/2018. 3, 4

[2] Hanahan D and Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; **144**: 646–674. 3

[3] National Cancer Institute. End-of-life care for people who have cancer. Available from: https://www.cancer.gov/about-cancer/advanced-cancer/care-choices/care-fact-sheet, accessed on 10/07/2018. 3

[4] IARC. Breast cancer screening: IARC Handbook of cancer prevention volume 15. *Lyon, France: International Agency for Research on Cancer* 2016; ISBN 978-92-832-3015-1. 3

[5] Engholm G, Ferlay J, Christensen N, Hansen HL, Hertzum-Larsen R, Johannesen TB, et al. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 8.1. *Association of the Nordic Cancer Registries. Danish Cancer Society* 2018; available from: http://www.ancr.nu, accessed on 05/07/2018. 3, 5, 6

[6] Olsson S, Andersson I, Karlberg I, Bjurstam N, Frodis E and Håkansson S. Implementation of service screening with mammography in Sweden: from pilot study to nationwide programme. *J Med Screen* 2000; **7**: 14–18. 3

[7] Dean PB and Pamilo M. Screening mammography in Finland–1.5 million examinations with 97 percent specificity. Mammography Working Group, Radiological Society of Finland. *Acta Oncol* 1999; **38**: 47–54. 3

[8] Sigurdsson K and Ólafsdóttir EJ. Population-based service mammography screening: the Icelandic experience. *Breast Cancer (Dove Med Press)* 2013; **5**: 17–25. 3

[9] Jørgensen KJ, Zahl PH and Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Womens Health* 2009; **9**: 36. 3

[10] Lynge E, Bak M, von Euler-Chelpin M, Kroman N, Lernevall A, Mogensen NB, et al. Outcome of breast cancer screening in Denmark. *BMC Cancer* 2017; **17**: 897. 3

[11] Kalager M, Tamimi RM, Bretthauer M and Adami HO. Prognosis in women with interval breast cancer: population based observational cohort study. *BMJ* 2012; **345**: e7536. 3

[12] Bergman O, Fredholm L, Hont G, Johansson E, Ljungman P, Munck-Wikland E, et al. Cancer i siffor – Populärvetenskapliga fakta om cancer. *Socialstyrelsen and Cancerfonden* 2018; ISBN 978-91-88161-18-5. 3, 8

[13] Vinnicombe SJ. Breast density: why all the fuss? *Clin Radiol* 2018; **73**: 334–357. 6

[14] Javed A and Lteif A. Development of the human breast. In: Seminars in plastic surgery. *Thieme Medical Publishers* 2013; **27**: 5. 6

[15] Radisky DC and Hartmann LC. Mammary involution and breast cancer risk: transgenic models and clinical studies. *J Mammary Gland Biol Neoplasia* 2009; **14**: 181–191. 6

[16] Lakhani SR, Ellis IO, Schnitt SJ, Tan PH and Van de Vijver MJ. WHO classification of tumours. *IARC WHO Classification of Tumours* 2012; **4**. 6

[17] Cancer Research UK. Cancer grading. Available from: https://www.cancerresearchuk.org/about-cancer/what-is-cancer/cancer-grading, accessed on 13/07/2018. 6

[18] AJCC. AJCC Cancer Staging Manual, 8th Ed. *Springer* 2017; ISBN 978-3-319-40617-6. 6, 7

[19] Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* 2017; **67**: 93–99. 7

[20] Hofvind S, Tsuruda K, Mangerud G, Ertzaas AK, Holen ÅS, Pedersen K, et al. The Norwegian Breast Cancer Screening Program, 1996-2016: celebrating 20 years of organised mammographic screening. *Cancer in Norway* 2017. 7

[21] Saadatmand S, Bretveld R, Siesling S and Tilanus-Linthorst MM. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173 797 patients. *BMJ* 2015; **351**: h4901. 7

[22] Sigurdsson H, Baldetorp B, Borg Å, Dalberg M, Fernö M, Killander D, et al. Indicators of prognosis in node-negative breast cancer. *NEJM* 1990; **322**: 1045–1053. 7

[23] Hedley DW, Rugg CA and Gelber RD. Association of DNA index and S-phase fraction with prognosis of nodes positive early breast cancer. *Cancer Res* 1987; **47**: 4729–4735. 7

[24] Gerdes J, Li L, Schlueter C, Duchrow M, Wohlenberg C, Gerlach C, et al. Immunobiochemical and molecular biologic characterization of the cell proliferation-associated nuclear antigen that is defined by monoclonal antibody Ki-67. *Am J Pathol* 1991; **138**: 867–873. 7

[25] van Dierendonck JH, Keijzer R, van de Velde CJH and Cornelisse CJ. Nuclear distribution of the Ki-67 antigen during the cell cycle: comparison with growth fraction in human breast cancer cells. *Cancer Res* 1989; **49**: 2999–3006. 7

[26] Lopez F, Belloc F, Lacombe F, Dumain P, Reiffers J, Bernard P, et al. Modalities of synthesis of Ki67 antigen during the stimulation of lymphocytes. *Cytometry* 1991; **12**: 42–49. 7

[27] Urruticoechea A, Smith IE and Dowsett M. Proliferation marker Ki-67 in early breast cancer. *J Clin Oncol* 2005; **23**: 7212–7220. 7

[28] Giuliano AE, Edge SB and Hortobagyi GN. Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. *Ann Surg Oncol* 2018; **25**: 1783–1785. 7

[29] Bregni G, Meneghini E, Galli G, Cavalieri S, Di Salvo F, Amash H, et al. Breast cancer Ki67, tumor size and axillary nodes relationship: it's complicated. *Ann Oncol* 2016; **27**: suppl_6. 7

[30] Harris JR, Lippman ME, Osborne CK and Morrow M. Diseases of the Breast. *Lippincott Williams & Wilkins* 2012. 8

[31] Singh D, Malila N, Pokhrel A and Anttila A. Association of symptoms and breast cancer in population-based mammography screening in Finland. *Int J Cancer* 2015; **136**: E630–E637. 8

[32] Haakinson DJ, Leeds SG, Dueck AC, Gray RJ, Wasif N, Stucky CCH, et al. The impact of obesity on breast cancer: a retrospective review. *Ann Surg Oncol* 2012; **19**: 3012–3018. 8, 9, 10, 20, 52

[33] Biglia N, Peano E, Sgandurra P, Moggio G, Pecchio S, Maggiorotto F, et al. Body mass index (BMI) and breast cancer: impact on tumor histopathologic features, cancer subtypes and recurrence rate in pre and postmenopausal women. *Gynecol Endocrinol* 2013; **29**: 263–267. 8, 9.

[34] Cui Y, Whiteman MK, Flaws JA, Langenberg P, Tkaczuk KH and Bush TL. Body mass and stage of breast cancer at diagnosis. *Int J Cancer* 2002; **98**: 279–283. 8, 9, 10, 20, 52

[35] Oestreicher N, White E, Lehman CD, Mandelson MT, Porter PL and Taplin SH. Predictors of sensitivity of clinical breast examination (CBE). *Breast Cancer Res Treat* 2002; **76**: 73–81. 8

[36] Duffy SW, Day NE, Tabár L, Chen HH and Smith TC. Markov models of breast tumor progression: some age-specific results. *J Natl Cancer Inst Monogr* 1997; 93–97. 8, 21, 22

[37] Aebi S, Gelber S, Castiglione-Gertsch M, Gelber RD, Collins J, Thürlimann B, et al. Is chemotherapy alone adequate for young women with oestrogen-receptor-positive breast cancer? *Lancet* 2000; **355**: 1869–1874. 8.

[38] Kopans DB, Rafferty E, Georgian-Smith D, Yeh E, D'Alessandro H, Moore R, et al. A simple model of breast carcinoma growth may provide explanations for observations of apparently complex phenomena. *Cancer* 2003; **97**: 2951–2959. 8

[39] Cancer Research UK. Breast cancer incidence (invasive) statistics. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#heading-One, accessed on 10/07/2018. 8

[40] Eriksson L, Czene K, Rosenberg L, Humphreys K and Hall P. The influence of mammographic density on breast tumor characteristics. *Breast Cancer Res Treat* 2012; **134**: 859–866. 9, 20, 31, 34

[41] Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. *NEJM* 2007; **356**: 227–236. 9, 20

[42] Duffy SW, Morrish OWE, Allgood PC, Black R, Gillan MGC, Willsher P, et al. Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer. *Eur J Cancer* 2018; **88**: 48–56. 9

[43] McCormack VA and dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 1159–1169. 9

[44] Hunter DJ and Willett WC. Diet, body size, and breast cancer. *Epidemiol Rev* 1993; **15**: 110–132. 9, 10

[45] Michels KB, Terry KL and Willett WC. Longitudinal study on the role of body size in premenopausal breast cancer. *Arch Intern Med* 2006; **166**: 2395–2402. 9

[46] American Cancer Society. Breast cancer facts & figures 2017–2018. *Atlanta: American Cancer Society, Inc* 2017. 10

[47] Boyd NF, Martin LJ, Sun L, Guo H, Chiarelli A, Hislop G, et al. Body size, mammographic density, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 2086–2092. 10

[48] Strand F, Humphreys K, Holm J, Eriksson M, Törnberg S, Hall P, et al. Long-term prognostic implications of risk factors associated with tumor size: a case study of women regularly attending screening. *Breast Cancer Res* 2018; **20**: 31. 10, 20, 52

[49] Feigelson HS, Jonas CR, Teras LR, Thun MJ and Calle EE. Weight gain, body mass index, hormone replacement therapy, and postmenopausal breast cancer in a large prospective study. *Cancer Epidemiol Biomarkers Prev* 2004; **13**: 220–224. 10

[50] Greendale GA, Reboussin BA, Slone S, Wasilauskas C, Pike MC and Ursin G. Postmenopausal hormone therapy and change in mammographic density. *J Natl Cancer Inst* 2003; **95**: 30–37. 10

[51] van Seumeren I. Weight gain and hormone replacement therapy: are women's fears justified? *Maturitas* 2000; **34**: S3–S8. 10

[52] Huang Z, Hankinson SE, Colditz GA, Stampfer MJ, Hunter DJ, Manson JE, et al. Dual effects of weight and weight gain on breast cancer risk. *JAMA* 1997; **278**: 1407–1411. 10

[53] Clamp A, Danson S and Clemons M. Hormonal and genetic risk factors for breast cancer. *Surgeon* 2003; **1**: 23–31. 10

[54] Press DJ, Sullivan-Halley J, Ursin G, Deapen D, McDonald JA, Strom BL, et al. Breast cancer risk and ovariectomy, hysterectomy, and tubal sterilization in the women's contraceptive and reproductive experiences study. *Am J Epidemiol* 2010; **173**: 38–47. 10

[55] Rebbeck TR, Friebel T, Lynch HT, Neuhausen SL, van't Veer L, Garber JE, et al. Bilateral prophylactic mastectomy reduces breast cancer risk in BRCA1 and BRCA2 mutation carriers: the PROSE Study Group. *J Clin Oncol* 2004; **22**: 1055–1062. 10

[56] Anderson WF and Matsuno R. Breast cancer heterogeneity: a mixture of at least two main types? *J Natl Cancer Inst* 2006; **98**: 948–951. 10

[57] Boyd N, Martin L, Stone J, Little L, Minkin S and Yaffe M. A longitudinal study of the effects of menopause on mammographic features. *Cancer Epidemiol Biomarkers Prev* 2002; **11**: 1048–1053. 10

[58] Perry N, Broeders M, de Wolf C, Törnberg S, Holland R and von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. —summary document. *Annal Oncol* 2008; **19**: 614–622. 11, 12

[59] Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG and Wilcox M. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013; **108**: 2205–2240. 11, 12, 20

[60] Gøtzsche PC and Nielsen M. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2011; CD001877. 11

[61] Gøtzsche PC and Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2013; CD001877. 11

[62] Holm J. Aggressive Breast Cancer: Epidemiological Studies Addressing Disease Heterogeneity. PhD Thesis. *Karolinska Institutet* 2018. 11

[63] Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *NEJM* 2005; **353**: 1784–1792. 11, 55

[64] Alagoz O, Berry DA, de Koning HJ, Feuer EJ, Lee SJ, Plevritis SK, et al. Introduction to the Cancer Intervention and Surveillance Modeling Network (CISNET) breast cancer models. *Med Decis Making* 2018; **38**: 3S–8S. 11, 26

[65] Plevritis SK, Munoz D, Kurian AW, Stout NK, Alagoz O, Near AM, et al. Association of screening and treatment with breast cancer mortality by molecular subtype in US women, 2000-2012. *JAMA* 2018; **319**: 154–164. 11, 26, 55

[66] Esserman L and O'Kane ME. Moving beyond the breast cancer screening debate. *J Womens Health (Larchmt)* 2014; **23**: 629–630. 12

[67] Shieh Y, Eklund M, Madlensky L, Sawyer SD, Thompson CK, Stover F, e al. Breast cancer screening in the precision medicine era: risk-based screening in a population-based trial. *J Natl Cancer Inst* 2017; **109**: djw290. 12

[68] Wilkinson L, Thomas V and Sharma N. Microcalcification on mammography: approaches to interpretation and biopsy. *Br J Radiol* 2016; **90**: 20160594. 12

[69] Kim SJ, Moon WK, Cho N, Cha JH, Kim SM and Im JG. Computer-aided detection in digital mammography: comparison of craniocaudal, mediolateral oblique, and mediolateral views. *Radiology* 2006; **241**: 695–701. 12

[70] Berg WA, Gutierrez L, NessAiver MS, Carter WB, Bhargavan M, Lewis RS, et al. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology* 2004; **233**: 830–849. 12, 56

[71] D'Orsi CJ. ACR BI-RADS atlas: breast imaging reporting and data system. *American College of Radiology* 2013. 12

[72] Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer* 1976; **37**: 2486–2492. 12

[73] Gram IT, Funkhouser E and Tabár L. The Tabar classification of mammographic parenchymal patterns. *Eur J Radiol* 1997; **24**: 131–136. 12

[74] Byng JW, Boyd NF, Fishell E, Jong RA and Yaffe MJ. The quantitative analysis of mammographic densities. *Phys Med Biol* 1994; **39**: 1629–1638. 13

[75] Li J, Szekely L, Eriksson L, Heddson B, Sundbom A, Czene K, et al. High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. *Breast Cancer Res* 2012; **14**: R114. 13, 34

[76] Strand F, Humphreys K, Cheddad A, Törnberg S, Azavedo E, Shepherd J, et al. Novel mammographic image features differentiate between interval and screen-detected breast cancer: a case-case study. *Breast Cancer Res* 2016; **18**: 100. 13, 20, 46

[77] Cheddad A, Czene K, Eriksson M, Li J, Easton D, Hall P, et al. Area and volumetric density estimation in processed full-field digital mammograms for risk assessment of breast cancer. *PLoS One* 2014; **9**: e110690. 13, 35

[78] Zheng Y, Keller BM, Ray S, Wang Y, Conant EF, Gee JC, et al. Parenchymal texture analysis in digital mammography: a fully automated pipeline for breast cancer risk assessment. *Med Phys* 2015; **42**: 4149–4160. 13

[79] Cox B and Sneyd MJ. Bias in breast cancer research in the screening era. *Breast* 2013; **22**: 1041–1045. 14, 18, 19

[80] Holm J, Humphreys K, Li J, Ploner A, Cheddad A, Eriksson M, et al. Risk factors and tumor characteristics of interval cancers by mammographic density. *J Clin Oncol* 2015; **33**: 1030–1037. 14, 15, 20

[81] Aro AR, De Koning HJ, Absetz P and Schreck M. Two distinct groups of non-attenders in an organized mammography screening program. *Breast Cancer Res Treat* 2001; **70**: 145–153. 15

[82] Domingo L, Salas D, Zubizarreta R, Baré M, Sarriugarte G, Barata T, et al. Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain. *Breast Cancer Res* 2014; **16**: R3. 15

[83] Yen AMF, Wu WYY, Tabar L, Duffy SW, Smith RA and Chen HH. Initiators and promoters for the occurrence of screen-detected breast cancer and the progression to clinically-detected interval breast cancer. *J Epidemiol* 2017; **27**: 98–106. 16, 56

[84] Eriksson L, Czene K, Rosenberg L, Humphreys K and Hall Per. Possible influence of mammographic density on local and locoregional recurrence of breast cancer. *Breast Cancer Res* 2013; **15**: R56. 19

[85] Lee SJ, Li X, Huang H and Zelen M. The Dana-Farber CISNET model for breast cancer screening strategies: An update. *Med Decis Making* 2018; **38**: 44S–53S. 20, 21, 22, 27, 55, 56

[86] Esserman LJ, Thompson IM, Reid B, Nelson P, Ransohoff DF, Welch HG, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *Lancet Oncol* 2014; **15**: e234–e242. 20

[87] Plevritis SK, Salzman P, Sigal BM and Glynn PW. A natural history model of stage progression applied to breast cancer. *Stat Med* 2007; **26**: 581–595. 21, 24, 25, 27, 37, 43, 57

[88] Duffy SW, Chen HH, Tabar L and Day NE. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. *Stat Med* 1995; **14**: 1531–1543. 21, 22

[89] Tan KHX, Simonella L, Wee HL, Roellin A, Lim YW, Lim WY, et al. Quantifying the natural history of breast cancer. *Br J Cancer* 2013; **109**: 2035–2043. 22, 56

[90] Taghipour S, Banjevic D, Miller AB, Montgomery N, Jardine AKS and Harvey BJ. Parameter estimates for invasive breast cancer progression in the Canadian National Breast Screening Study. *Br J Cancer* 2013; **108**: 542–548. 21, 22, 57

[91] Uhry Z, Hédelin G, Colonna M, Asselain B, Arveux P, Rogel A, et al. Multi-state Markov models in cancer screening evaluation: a brief review and case study. *Stat Methods Med Res* 2010; **19**: 463–486. 21, 22

[92] Weedon-Fekjær H, Tretli S and Aalen OO. Estimating screening test sensitivity and tumour progression using tumour size and time since previous screening. *Stat Methods Med Res* 2010; **19**: 507–527. 21, 22, 25, 37, 40, 42, 45, 46, 49, 56

[93] Stirzaker D. Stochastic processes and models. *Oxford University Press* 2005. 21

[94] Prevost TC, Launoy G, Duffy SW and Chen HH. Estimating sensitivity and sojourn time in screening for colorectal cancer: a comparison of statistical approaches. *Am J Epidemiol* 1998; **148**: 609–619. 22

[95] de Koning HJ, Draisma G, Fracheboud J and de Bruijn A. Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. *Breast Cancer Res* 2006; **8**: 202. 22

[96] Duffy SW, Nagtegaal ID, Wallis M, Cafferty FH, Houssami N, Warwick J, et al. Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *Am J Epidemiol* 2008; **168**: 98–104. 22, 23, 46, 57

[97] Talkington A and Durrett R. Estimating tumor growth rates in vivo. *Bull Math Biol* 2015; **77**: 1934–1954. 23, 24

[98] Heuser L, Spratt JS and Polk JHC. Growth rates of primary breast cancers. *Cancer* 1979; **43**: 1888–1894. 24

[99] Spratt JA, Von Fournier D, Spratt JS and Weber EE. Decelerating growth and human breast cancer. *Cancer* 1993; **71**: 2013–2019. 24, 25

[100] Hartung N, Mollard S, Barbolosi D, Benabdallah A, Chapuisat G, Henry G, et al. Mathematical modeling of tumor growth and metastatic spreading: validation in tumor-bearing mice. *Cancer Res* 2014; **74**: 6397–6407. 24

[101] Atkinson EN, Bartoszyński R, Brown BW and Thompson JR. On estimating the growth function of tumors. *Math Biosc* 1983; **67**: 145–166. 24, 25, 37

[102] Brown BW, Atkinson EN, Bartoszynski R, Thompson JR and Montague ED. Estimation of human tumor growth rate from distribution of tumor size at detection. *J Natl Cancer Inst* 1984; **72**: 31–38. 24, 25, 37

[103] Klein JP and Bartoszynski R. Estimation of growth and metastatic rates of primary breast cancer. *Math Pop Dyn* 1991; 397–412. 24

[104] Bartoszyński R, Edler L, Hanin L, Kopp-Schneider A, Pavlova L, Tsodikov A, et al. Modeling cancer detection: tumor size as a source of information on unobservable stages of carcinogenesis. *Math Biosc* 2001; **171**: 113–142. 24, 37

[105] Chia YL, Salzman P, Plevritis SK and Glynn PW. Simulation-based parameter estimation for complex models: a breast cancer natural history modelling illustration. *Stat Methods Med Res* 2004; **13**: 507–524. 24, 26

[106] Weedon-Fekjær H, Lindqvist BH, Vatten LJ, Aalen OO and Tretli S. Breast cancer tumor growth estimated through mammography screening data. *Breast Cancer Res* 2008; **10**: R41. 25, 37, 40, 42, 45, 49, 56

[107] Hanin LG and Yakovlev AY. Multivariate distributions of clinical covariates at the time of cancer detection. *Stat Methods Med Res* 2004; **13**: 457–489. 25

[108] Hanin LG, Tsodikov AD and Yakovlev AY. Optimal schedules of cancer surveillance and tumor size at detection. *Math Comput Model* 2001; **33**: 1419–1430. 25

[109] Stout NK, Knudsen AB, Kong CY, McMahon PM and Gazelle GS. Calibration methods used in cancer simulation models and suggested reporting guidelines. *Pharmacoeconomics* 2009; **27**: 533–545. 26

[110] Tan SYGL, Van Oortmarssen GJ, De Koning HJ, Boer R and Habbema JDF. Chapter 9: the MISCAN-Fadia continuous tumor growth model for breast cancer. *J Natl Cancer Inst Monogr* 2006; 56–65. 26, 27, 55

[111] Munoz D, Near AM, Van Ravesteyn NT, Lee SJ, Schechter CB, Alagoz O, et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast cancer mortality. *J Natl Cancer Inst* 2014; **106**. 27

[112] Munoz DF, Xu C and Plevritis SK. A Molecular Subtype–Specific Stochastic Simulation Model of US Breast Cancer Incidence, Survival, and Mortality Trends from 1975 to 2010. *Med Decis Making* 2018; **38**: 89S–98S. 27

[113] Plevritis SK, Sigal BM, Salzman P, Rosenberg J and Glynn P. Chapter 12: a stochastic simulation model of US breast cancer mortality trends from 1975 to 2000. *J Natl Cancer Inst Monogr* 2006; 86–95. 27

[114] Shwartz M. A mathematical model used to analyze breast cancer screening strategies. *Oper Res* 1978; **26**: 937–955. 27

[115] Clarke LD, Plevritis SK, Boer R, Cronin KA and Feuer EJ. Chapter 13: A Comparative Review of CISNET Breast Models Used To Analyze US Breast Cancer Incidence and Mortality Trends. *J Natl Cancer Inst Monogr* 2006; 96–105. 27

[116] National Cancer Institute, Cancer Intervention and Surveillance Modeling Network. Breast Cancer Modeling. Available from: https://cisnet.cancer.gov/breast/, accessed on 26/07/2018. 27

[117] Barlow L, Westergren K, Holmberg L and Talbäck M. The completeness of the Swedish Cancer Register–a sample survey for year 1998. *Acta Oncol* 2009; **48**: 27–33. 31

[118] Ludvigsson JF, Otterblad-Olausson P, Pettersson BU and Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 2009; **24**: 659–667. 31

[119] Rosenberg LU, Magnusson C, Lindström E, Wedrén S, Hall P, and Dickman PW. Menopausal hormone therapy and other breast cancer risk factors in relation to the risk of different histological subtypes of breast cancer: a case-control study. *Breast Cancer Res* 2006; **8**: R11. 31

[120] Fasching PA, Gass P and Hein A. Neoadjuvant Treatment of Breast Cancer-Advances and Limitations. *Breast Care* 2016; **11**: 313–314. 32

[121] Pawitan Y. In all likelihood: statistical modelling and inference using likelihood. *Oxford University Press* 2001. 39

[122] Held L and Sabanés Bové D. Applied statistical inference. *Springer, Berlin, Heidelberg* 2014. 39

[123] Byrd RH, Lu P, Nocedal J, and Zhu C. A limited memory algorithm for bound constrained optimization. *SIAM J Sci Comput* 1995; **16**: 1190–1208. 40

[124] R Core Team. R: A Language and Environment for Statistical Computing. *Vienna, Austria, R Foundation for Statistical Computing* 2017. Available from: https://www.R-project.org/. 40, 47

[125] Broyden CG. The convergence of a class of double-rank minimization algorithms: 2. The new algorithm. *IMA J Appl Math* 1970; **6**: 222–231. 40

[126] Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals Math Stat* 1938; **9**: 60–62. 41

[127] Cole SR, Chu H and Greenland S. Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *Am J Epidemiol* 2013; **179**: 252–260. 41

[128] Conover WJ. Practical nonparametric statistics, 3rd ed. *Wiley New York* 1999. 41

[129] Isheden G and Humphreys K. Modelling breast cancer tumour growth for a stable disease population. *Stat Methods Med Res* 2017. Epub ahead of print. 42, 44, 47, 57

[130] Habbema JDF, Van Oortmarssen GJ, Lubbe JTN and Van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. *Comput Methods Programs Biomed* 1985; **20**: 79–93. 55

[131] Chiu SY, Duffy S, Yen AM, Tabár L, Smith RA and Chen HH. Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. *Cancer Epidemiol Biomarkers Prev* 2010; **19**: 1219–1228. 56

[132] Cong XJ, Shen Y and Miller AB. Estimation of age-specific sensitivity and sojourn time in breast cancer screening studies. *Stat Med* 2005; **24**: 3123–3138. 56

[133] Shen Y and Zelen M. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. *J Clin Oncol* 2001; **19**: 3490–3499. 57

[134] Weedon-Fekjær H, Vatten LJ, Aalen OO, Lindqvist B and Tretli S. Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening: new results. *J Med Screen* 2005; **12**: 172–178. 57

[135] Chen Y, Brock G and Wu D. Estimating key parameters in periodic breast cancer screening—application to the Canadian national breast screening study data. *Cancer Epidemiol* 2010; **34**: 429–433. 57

[136] Esserman LJ, WISDOM Study and Athena Investigators. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer* 2017; **3**: 34. 61

[137] Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018; **359**: 926–930. 62